KOMBINASI METODE SMOTE DAN PCA UNTUK MENINGKATKAN PERFORMA KLASIFIKASI FIBROSIS HATI DATA PASIEN HEPATITIS C BERBASIS *EXTREME LEARNING MACHINE*

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh :

RANDA WAHYU SAPUTRA

H76216072

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2021

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini,

NAMA : Randa Wahyu Saputra
NIM : H76216072
Program Studi : Sistem Informasi
Angkatan : 2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penelitian skripsi saya yang berjudul "KOMBINASI METODE SMOTE DAN PCA UNTUK MENINGKATKAN PERFORMA KLASIFIKASI FIBROSIS HATI DATA PASIEN HEPATITIS C BERBASIS *EXTREME LEARNING MACHINE*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian saya buat dengan sebenar-benarnya.

Surabaya, 26 Januari 2021
Yang Menyatakan,

METERAI TEMPEL
TGL 20
A94CAAHF858892885
6000
ENAM RIBU RUPIAH

(Randa Wahyu Saputra)
NIM : H76216072

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA	:	RANDA WAHYU SAPUTRA
NIM	:	H76216072
JUDUL	:	KOMBINASI METODE SMOTE DAN PCA UNTUK MENINGKATKAN PERFORMA KLASIFIKASI FIBROSIS HATI DATA PASIEN HEPATITIS C BERBASIS *EXTREME LEARNING MACHINE*

Ini telah diperiksa dan disetujui untuk diujikan

Surabaya, 14 Desember 2020

Dosen Pembimbing 1	Dosen Pembimbing 2
Dwi Rolliawati, MT	Khalid, M. Kom
NIP. 197909272014032001	NIP. 197906092014031002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Randa Wahyu Saputra ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 16 Desember 2020.

Mengesahkan,
Dewan Penguji

Penguji I	Penguji II
Dwi Rolliawati, MT	Khalid, M. Kom
NIP. 197909272014032001	NIP. 197906092014031002

Penguji III	Penguji IV
Mujib Ridwan, S.Kom., M.T	Indri Sudanawati Rozas, M.Kom
NIP. 198604272014031004	NIP. 198207212014032001

Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. Evi Fatimatur Rusydiyah, M.Ag
NIP. 197312272005012003

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama	: RANDA WAHYU SAPUTRA
NIM	: H76216072
Fakultas/Jurusan	: FAKULTAS SAINS DAN TEKNOLOGI / SISTEM INFORMASI
E-mail address	: wahyuranda16@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
■ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

KOMBINASI METODE SMOTE DAN PCA UNTUK MENINGKATKAN PERFORMA KLASIFIKASI FIBROSIS HATI DATA PASIEN HEPATITIS C BERBASIS *EXTREME LEARNING MACHINE*

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 26 Januari 2021

Penulis

(Randa Wahyu Saputra)
nama terang dan tanda tangan

ABSTRAK

KOMBINASI METODE SMOTE DAN PCA UNTUK MENINGKATKAN PERFORMA KLASIFIKASI FIBROSIS HATI DATA PASIEN HEPATITIS C BERBASIS *EXTREME LEARNING MACHINE*

Oleh :

Randa Wahyu Saputra

Hepatitis C merupakan infeksi penyakit yang disebabkan oleh Hepatitis C Virus dan menyerang fungsi organ hati. Kondisi tersebut memungkinkan terjadinya radang hati akibat sel-sel hati yang rusak atau biasa dikenal dengan istilah fibrosis hati. Hepatitis C hampir tidak memiliki gejala awal dan memiliki masa pengobatan terbilang cukup lama yaitu 7-26 minggu. Oleh sebab itu perlu dilakukan diagnosa awal untuk penyakit ini salah satunya yaitu dengan biopsi hati. Biopsi hati merupakan standar terbaik untuk diagnosis perkembangan fibrosis hati, namun cara ini cukup mahal dan kurang nyaman untuk pasien, sehingga dibuatlah sistem klasifikasi untuk memprediksi fibrosis hati untuk membantu dalam melakukan diagnosa penyakit Hepatits C. Dengan menggunakan kombinasi metode seleksi fitur PCA dan *oversampling* SMOTE, penelitian ini membandingkan dua *dataset* Hepatitis C yaitu *Hepatitis C Virus (HCV) for Egyptian patients Data Set* dan *HCV data Data Set* yang berasal dari UCI Repository untuk diklasifikasikan menggunakan metode *Extreme Learning Machine*. Kombinasi metode SMOTE dan PCA mampu meningkatkan nilai evaluasi *HCV data Data Set* dengan nilai awal *accuracy*, *sensitivity*, dan *specificity* berturut-turut adalah 90%, 44%, 47% semuanya meningkat menjadi 99%. sedangkan pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* nilai *accuracy, sensitivity,* dan *specificity* yang awalnya sama yaitu 26%, meningkat menjadi 31%.

Kata kunci: Hepatitis C, Fibrosis Hati, Exteme Learning Machine, PCA, SMOTE

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

COMBINATION OF SMOTE AND PCA METHODS TO IMPROVE THE LIVER FIBROSIS CLASSIFICATION PERFORMANCE OF HEPATITIS C PATIENTS BASED ON EXTREME LEARNING MACHINE

By :

Randa Wahyu Saputra

Hepatitis C is a disease infection caused by Hepatitis C virus and attacks the function of the liver. This condition allows inflammation of the liver due to damaged liver cells or commonly known as liver fibrosis. Hepatitis C has almost no initial symptoms and has a fairly long treatment period of 7-26 weeks. Therefore, it is necessary to make an early diagnosis for this disease, one of which is a liver biopsy. Liver biopsy is the best standard for the diagnosis of liver fibrosis progression, but this method is quite expensive and inconvenient for patients, so a classification system is created to predict liver fibrosis to assist in diagnosing liver disease C. Using a combination of PCA feature selection methods and SMOTE oversampling, This study compared two Hepatitis C datasets, namely the Hepatitis C Virus (HCV) for Egyptian patients Data Set and the HCV Data Set from the UCI Repository to be classified using the Extreme Learning Machine method. The combination of the SMOTE and PCA methods was able to increase the HCV evaluation value. Set the initial values for accuracy, sensitivity, and specificity were 90%, 44%, 47%, which of all increased to 99%. whereas in Hepatitis C Virus (HCV) for Egyptian patients Data Set the accuracy, sensitivity, and specificity were initially have the same values, namely 26%, increased to 31%.

Keywords: Hepatitis C, Liver Fibrosis, Extreme Learning Machine, PCA, SMOTE

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN i
LEMBAR PERSETUJUAN PEMBIMBING ii
PENGESAHAN TIM PENGUJI SKRIPSI iii
LEMBAR PERSETUJUAN PUBLIKASI ILMIAH iv
ABSTRAK v
ABSTRACT vi
DAFTAR ISI vii
DAFTAR TABEL ix
DAFTAR GAMBAR xi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 4
BAB II TINJAUAN PUSTAKA 6
2.1 Tinjauan Penelitian Terdauhulu 6
2.2 Virus Hepatitis C (HCV) 9
2.3 Infeksi HCV 9
2.4 Fibrosis Hati 9
2.5 Data Mining 10
2.6 Klasifikasi 11
2.7 K-fold cross validation 12
2.8 Principal Component Analysis 13
2.9 Extreme Learning Machine 14
2.9.1 Feed Forward Algorithm 15
2.9.2 Single Hidden Layer Feed Forward Neural Network (SLFNs) 16
2.9.3 Algoritma ELM 18
2.9.4 Struktur Jaringan Syaraf Tiruan ELM 19
2.10 Python 20
2.11 Confusion Matrix 20

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.12 Synthetic Minority Oversampling Technique 22
2.13 Exploratory Data Analysis 22
2.14 Integrasi Keilmuan 22
BAB III METODOLOGI PENELITIAN 25
3.1 Tahapan Penelitian..... 25
3.1.1 Perumusan Masalah 25
3.1.2 Studi Literatur 26
3.1.3 Pengumpulan Data 26
3.1.4 Pengolahan Data 31
3.1.5 Analisa Data..... 36
BAB IV HASIL DAN PEMBAHASAN 37
4.1. Hasil Klasifikasi ELM..... 37
4.1.1 Hepatitis C Virus (HCV) for Egyptian patients Data Set 37
4.1.1.1 Pre-Processing 37
4.1.1.2 Klasifikasi 38
4.1.1.3 Evaluating..... 38
4.1.2 HCV data Data Set..... 39
4.1.2.1 Pre-Processing 39
4.1.2.2 Klasfikasi 40
4.1.2.3 Evaluating..... 40
4.2. Hasil Klasifikasi ELM Dengan SMOTE *Oversampling* 41
4.2.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set* 41
4.2.2 *HCV data Data* 43
4.3. Hasil Klasifikasi ELM Dengan *Principal Component Analysis* 45
4.3.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set* 45
4.3.2 *HCV data Data Set*..... 49
4.4. Hasil Klasifikasi ELM Dengan Kombinasi SMOTE dan PCA..... 52
4.4.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set* 52
4.4.2 *HCV data Data Set*..... 56
4.5. Analisis Hasil 58
BAB V Kesimpulan dan Saran 64
5.1. Kesimpulan..... 64
5.2. Saran..... 65
DAFTAR PUSTAKA..... 66

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu 6
Tabel 3. 1 Perbandingan Dataset 30
Tabel 3. 2 Skenario Klasifikasi ELM 34
Tabel 4. 1 Hepatitis C Virus (HCV) for Egyptian patients Data Set Setelah Pre-processing dengan Min Max Scaler 37
Tabel 4. 2 Hasil klasifikasi ELM Hepatitis C Virus (HCV) for Egyptian patients Data Set 39
Tabel 4. 3 HCV data Data Set setelah Pre-processing dengan Min Max Scaler .. 39
Tabel 4. 4 Hasil klasifikasi ELM HCV data Data Set 41
Tabel 4. 5 Hasil klasifikasi ELM dengan SMOTE Hepatitis C Virus (HCV) for Egyptian patients Data Set 43
Tabel 4. 6 Hasil Klasifikasi ELM Dengan SMOTE HCV data Data Set 44
Tabel 4. 7 Hasil Klasifikasi PCA+ELM Dengan 25 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 46
Tabel 4. 8 Hasil Klasifikasi PCA+ELM Dengan 20 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 47
Tabel 4. 9 Hasil Klasifikasi PCA+ELM Dengan 17 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 47
Tabel 4. 10 Hasil Klasifikasi PCA+ELM Dengan 15 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 48
Tabel 4. 11 Hasil Klasifikasi PCA+ELM Dengan 12 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 48
Tabel 4. 12 Hasil Klasifikasi PCA+ELM dengan 10 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 49
Tabel 4. 13 Hasil Klasifikasi PCA+ELM Dengan 10 n-component HCV data Data Set 50
Tabel 4. 14 Hasil Klasifikasi PCA+ELM dengan 8 n-component HCV data Data Set 51
Tabel 4. 15 Hasil Klasifikasi PCA+ELM Dengan 6 n-component HCV data Data Set 51
Tabel 4. 16 Hasil Klasifikasi PCA+ELM Dengan 5 n-component HCV data Data Set 52
Tabel 4. 17 Hasil Klasifikasi PCA+ELM Dengan 3 n-component HCV data Data Set 52
Tabel 4. 18 Hasil Kasifikasi SMOTE+PCA+ELM Dengan 25 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 53
Tabel 4. 19 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 20 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 53
Tabel 4. 20 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 17 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 54

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 21 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 15 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 54
Tabel 4. 22 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 12 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set 55
Tabel 4. 23 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 10 n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set dataset Mesir 55
Tabel 4. 24 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 10 n-component HCV data Data Set 56
Tabel 4. 25 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 8 n-component HCV data Data Set 57
Tabel 4. 26 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 6 n-component HCV data Data Set 57
Tabel 4. 27 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 5 n-component HCV data Data Set 58
Tabel 4. 28 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 8 n-component HCV data Data Set 58
Tabel 4. 29 Perbandingan Hasil Klasifikasi Metode ELM..... 59
Tabel 4. 30 Perbandingan Hasil Klasifikasi ELM dengan SMOTE..... 59
Tabel 4. 31 Perbandingan Hasil Klasifikasi ELM dengan PCA 60
Tabel 4. 32 Perbandingan Hasil Klasifikasi ELM Dengan Kombinasi SMOTE + PCA 61
Tabel 4. 33 Perbandingan Hasil klasifikasi Kedua Dataset..... 61

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

Gambar 2. 1 Tahapan *Data Mining* 10
Gambar 2. 2 Model 3-fold cross validation 12
Gambar 2. 3 Arsitektur Feedforward 16
Gambar 2. 4 Arsitektur Extreme Learning Machine 19
Gambar 2. 5 Confusion Matrix 21
Gambar 3. 1 Alur Penelitian 25
Gambar 3. 2 Flow Min-max Scaler 31
Gambar 3. 3 Flow SMOTE Oversampling 32
Gambar 3. 4 Flow Metode Principal Component Analysis 33
Gambar 3. 5 Flow Metode Extreme Learning Machine 35
Gambar 4. 1 Jumlah Atribut Hepatitis C Virus (HCV) for Egyptian patients Data Set Sebelum Implementasi SMOTE 42
Gambar 4. 2 Jumlah Atribut Hepatitis C Virus (HCV) for Egyptian patients Data Set Setelah Implementasi SMOTE 42
Gambar 4. 3 Jumlah Atribut HCV data Data Set Sebelum Implementasi SMOTE 43
Gambar 4. 4 Jumlah Atribut HCV data Data Set Setelah implementasi SMOTE44
Gambar 4. 5 Diagram Hasil PCA Hepatitis C Virus (HCV) for Egyptian patients Data Set 45
Gambar 4. 6 Diagram Hasil PCA HCV data Data Set 49

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB I
PENDAHULUAN

1.1 Latar Belakang

Hepatitis C merupakan infeksi penyakit yang menyerang organ hati karena sebuah virus yang menyerang sel-sel dan fungsi hati sehingga menyebabkan kerusakan, penyakit ini disebabkan oleh hepatitis C Virus (Brataatmadja, 2003). Hepatitis C hampir tidak memberikan tanda gejala awal, namun dapat merusak fungsi organ hati yang berperan sebagai penetral racun dan sistem pencernaan makanan yang menguraikan sari-sari makanan untuk kemudian disebarkan ke seluruh organ tubuh sebagai energi. (Septiani, 2017). Infeksi virus Hepatitis C dapat mengakibatkan komplikasi hati, termasuk fibrosis, sirosis, kanker hati dan menjadi masalah utama kesehatan dunia dimana penderita penyakit ini mencapai 150 -170 juta orang, dengan perkiraan kematian terkait 35000 orang / tahun (Cacoub & Comarmond, 2017).

Diagnosa medis sangat dibutuhkan bagi para penderita penyakit hepatitis C sebagai tindakan awal sebelum penyakit menjadi lebih parah (kronis), Hal ini dikarenakan masa pengobatan penyakit ini terbilang cukup lama yaitu 6 bulan atau 7 - 26 minggu (Brataatmadja, 2003). Sistem klasifikasi secara otomatis akan sangat bermanfaat dalam penanganan penyakit Hepatitis C (Fridayanthie, 2015). Berdasarkan observasi data pasien lain di masa lalu atau yang sudah terekap analisis dan komparasi metode klasifikasi data mining dapat digunakan untuk mengembangkan sistem diagnosis medis untuk memprediksi penyakit hepatitis c yaitu fibrosis hati (Khomsah, 2018). Fibrosis hati adalah akumulasi protein matriks ekstraseluler termasuk kolagen yang berlebihan dan sering terjadi pada jenis penyakit hati kronis (Bataller & Brenner, 2005). Melalui metode tersebut, kondisi pasien penderita hepatitis dapat diprediksi tingkatan fibrosis hatinya apakah normal atau kronis. Mempertimbangkan permasalahan tersebut, penelitian ini bermaksud mengimplementasikan metode klasifikasi *data mining* untuk prediksi fibrosis hati pada pasien terdampak Hepatitis C.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Klasifikasi merupakan cara pengelompokkan data berdasarkan karakteristik dan ciri-ciri data tersebut. Klasifikasi memiliki beberapa meode seperti *Decission Tree*, *Naïve Bayes*, *Neural Network*, *Support Vector Machine*, dan *Fuzzy* (Wibawa, 2018). Beberapa penelitian yang membandingkan metode klasifikasi ELM, SVM, ANN, PSO-SVM, *Backpropagation* (Wibawa, 2018), (Sreekanth, Rajesh, & Satheeshkumar, 2015), (Pratiwi & Harianto, 2019), menunjukan bahwa metode ELM memiliki *accuracy* tertinggi dengan rata rata 93% dengan tingkat *error* 14,84% paling rendah dibanding metode lainnya. Dari pemaparan tersebut metode ELM dipilih dengan alasan memiliki nilai *accuracy* tinggi dan *error* yang cukup rendah. Selain itu metode ELM dikenal sebagai metode jaringan saraf tiruan paling cepat dikarenakan berbasis *single hidden layer* (Hidayat, 2012).

Dari beberapa pemaparan penelitian terdahulu yang menggunakan metode ELM memiliki rekomendasi peningkatan kerja algoritma dengan diterapkannya metode *pre-processing* seperti *feature selection* (Ravikumar, 2016) untuk meningkatkan efisiensi model klasifikasi dalam hal *accuracy* dengan membuang fitur dan set fitur yang tidak perlu (Priyadarshini, Dash, & Mishra, 2014; Wu, Wang, Chen, & Zhao, 2016). Sehingga berdasarkan GAP tersebut implementasi metode *feature selection* pada metode klasifikasi ELM menjadi tujuan utama dari penelitian ini.

Pre-processing merupakan tahapan awal proses klasifikasi untuk mempersiapkan data mentah melalui *eliminasi, cleaning, dan transformasi* yang dilakukan sebelum proses klasifikasi agar mudah dieksekusi oleh sistem (Mujilahwati, 2016). Pada proses *pre-processing* metode *fitur seletion* memberikan pengaruh besar pada hasil klasifikasi, sebagaimana paparan riset-riset sebelumnya (Komputer, Matematika, Ilmu, & Alam, 2015), (G. A. Irawan & Muliantara, 2017), (Adiwijaya, 2018) menunjukkan metode *fitur selection Principal Component Analysis* (PCA) dapat meningkatkan rata-rata *accuracy* hingga 90% dan mengurangi selisih waktu komputasi setelah direduksi hingga 88,109 detik. Metode PCA terbukti efektif dalam mengurangi fitur data, dimana hasil proses metode ini merupakan fitur data pilihan yang representasinya telah optimal (Adiwijaya, 2018) sehingga digunakan pada penelitian ini.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Namun, pada penelitian medis keseimbangan data meberikan efek yang berpengaruh terhadap hasil klasifikasi. Seperti jika kelas negatif lebih banyak dan memiliki selisih yang cukup banyak dibandingkan kelas positif, maka hasil klasifikasi akan cenderung menghasilkan negatif meskipun *accuracy* yang didapatkan sangat tinggi (Ramezankhani et al., 2016). Ada beberapa solusi untuk menangani jumlah *dataset* yang tidak seimbang seperti menggunakan metode *oversampling*. *Synthetic Minority Oversampling Technique* (SMOTE) merupakan metode *oversampling* yang menghasilkan sampel baru dari kelas minoritas secara artifisial dengan menginterpolasi yang sudah ada sebelumnya (Maldonado, López, & Vairetti, 2019). Pada penelitian "Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection", SMOTE mampu meningkatkan *accuracy* hasil klasifikasi dari 69,43 % menjadi 82,62 % (Sreejith, Khanna Nehemiah, & Kannan, 2020).

Adapun penelitian mengenai klasifikasi data pasien terdampak hepatitis C akan membandingkan dua *dataset* berbeda *Hepatitis C Virus (HCV) for Egyptian patients Data Set* dan *HCV data Data Set* yang berasal dari UCI Repository. Kedua *dataset* memiliki jumlah atribut yang berbeda, namun memiliki beberapa jenis atribut yang sama. Kedua *dataset* akan diberikan sekenario yang sama dalam implementasi metode SMOTE dan PCA yang selanjutnya *dataset* tersebut akan dilakukan klasifikasi menggunakan algoritma *Extreme Learning Machine* yang berbasis *single hidden layer*. Perbandingan hasil klasifikasi dari dua *dataset* tersebut akan menentukan apakah metode PCA dan SMOTE memberikan pengaruh dalam peningkatan hasil evaluasi atau tidak pada klasifikasi menggunakan metode ELM.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, pada penelitian ini dapat dirumusukan permasalah sebagai berikut:

1. Bagaimana hasil klasifikasi metode *Extreme Learning Machine* pada klasifikasi data pasien Hepatitis C?
2. Bagaimana hasil performa metode SMOTE terhadap klasifikasi data pasien Hepatitis C berbasis *Extreme Learning Machine*?

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3. Bagaimana hasil performa metode *principal component analysis* terhadap klasifikasi data pasien Hepatitis C berbasis *Extreme Learning Machine*?
4. Bagaimana hasil performa kombinasi metode SMOTE *oversampling* dan *principal component analysis* terhadap klasifikasi data pasien Hepatitis C *Extreme Learning Machine*?

1.3 Batasan Masalah

Batasan masalah akan membuat penelitian ini lebih fokus, mengingat ruang lingkup penelitian yang begitu luas. Beberapa batasan masalah tersebut yaitu:

1. Pada penelitian ini menggunakan 2 *dataset* yaitu *Hepatitis C Virus (HCV) for Egyptian patients Data Set* (Learning, 2019) yang berasal dari Mesir dan *HCV data Data Set* yang berasal dari Jerman (Learning, 2020).
2. Aktivasi fungsi yang digunakan pada algoritma *Extreme Learning Machine* adalah *Sigmoid*

1.4 Tujuan Penelitian

Bersumber dari latar belakang dan rumusan masalah yang telah dipaparkan, dapat diperoleh tujuan penelitian sebagai berikut:

1. Dapat mengetahui hasil klasifikasi metode *Extreme Learning Machine* pada klasifikasi data pasien Hepatitis C.
2. Dapat mengetahui hasil performa metode SMOTE terhadap klasifikasi data pasien Hepatitis C berbasis *Extreme Learning Machine*.
3. Dapat mengetahui hasil performa metode *principal component analysis* terhadap klasifikasi data pasien Hepatitis C berbasis *Extreme Learning Machine.*
4. Dapat mengetahui hasil performa kombinasi metode SMOTE *oversampling* dan *principal component analysis* terhadap klasifikasi data pasien Hepatitis C *Extreme Learning Machine.*

1.5 Manfaat Penelitian

Tujuan penelitian yang akan dicapai pada penelitian ini diharapkan memberikan manfaat secara langsung maupun secara tidak langsung.

Adapun manfaat penelitian ini adalah:

1. Secara Akademis

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

a. Dapat menghasilkan sumbangan ilmiah khususnya dalam ilmu kesehatan dari inovasi dalam penggunaan metode *Extreme Learning Machine* untuk klasifikasi penyakit Hepatitis C.
b. Dapat digunakan sebagai referensi untuk penelitian di masa mendatang yang berkaitan dengan Hepatitis C.

2. Secara Aplikatif

   Memberikan manfaat melalui hasil penelitian tentang cara mengidentifikasi fibrosis hati sehingga dapat mengurangi adanya penyebaran penyakit yang lebih parah (kronis).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II
TINJAUAN PUSTAKA

2.1 Tinjauan Penelitian Terdauhulu

Berisi pemaparan penelitian terdahulu yang relevan sebagai rujukan dan landasan utama dalam melakukan penelitian ini. Beberapa penelitian tersebut adalah sebagai berikut:

Tabel 2. 1 Penelitian Terdahulu

No.	Kategori Penelitian	Hasil	Korelasi
1.	Perbandingan metode ELM dengan SVM dan ANN (Sreekanth et al., 2015)	Dalam penelitian ini *extreme learning machine* digunakan untuk klasifikasi badai dan curah hujan yang dibandingkan dengan SVM dan ANN. ELM memiliki hasil yang lebih baik dan cocok untuk peramalan cuaca.	Metode ELM dapat digunakan sebagai metode untuk melakukan klasifikasi berkaitan dengan *meteorology*.
2.	Hepatitis C (Cacoub & Comarmond, 2017)	Singkatnya, selain menyerang hati, infeksi kronis HCV sering mengarah atau berpotensi pada penyakit bidang *reumatologi* seperti *ligamen*, sendi, dan otot	Para ahli *reumatologi* atau dokter harus meningkatkan skrining infeksi HCV pada pasien mereka
3.	Klasifikasi ELM unutk data Medis (Utomo & Kardiana, 2014)	Kinerja ELM pada umumnya lebih baik daripada BP dalam diagnosis kanker payudara. Meskipun tingkat *specificity* sedikit lebih rendah dari BP, dapat dilihat dengan jelas bahwa ELM secara luar biasa meningkatkan tingkat *sensitivity* dan *accuracy*.	Pengembangan antarmuka dalam aplikasi seluler, desktop, atau web mungkin bermanfaat. Ketiga, ada kasus baru yang ditambahkan secara teratur di rumah sakit. Mengembangkan sistem diagnosis cerdas yang tidak hanya dapat belajar dari data yang tersedia di repositori tetapi juga dari data yang baru tersedia akan diperlukan.
4.	Klasifikasi ELM untuk data Medis (Xie, Li, & Ma, 2016)	Hasil percobaan menunjukkan bahwa ELM dalam proses pengenalan malignant masses dan benign menjadi	Studi lebih lanjut atau penelitian kedepannya dapat dilakukan dengan penekanan pada model klasifikasi yang lebih generatif

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		yang terbaik dibangingkan dengan metode pengklasifikasi populer lainnya yang ada sesuai dengan database DDSM dan MIAS	
5.	Kombinasi metode ELM dan PCA (Wiyono & Imah, 2018)	Hasil testing menunjukkan *accuracy* tertinggi terdapat pada percobaan dengan menggunakan 90 fitur dan fungsi aktivasi *sigmoid*, dengan rasio data train : tes = 4 : 1, dengan *accuracy* sebesar 715 (71.5%) , dengan waktu training selama 062 detik dan waktu testing selama 015 detik	Kedua algoritma tersebut dapat diimplementasikan untuk *dataset* lain dengan jumlah yang lebih banyak dari jumlah *dataset* yang digunakan saat ini.
6.	Kombinasi metode ELM dan PCA (Izzuddin, Wahyudi, & Kartini, 2018)	Hasil Implementasi metode PCA dan ELM dapat membedakan tanaman gulma dengan padi (Oryza sativa L) dalam kasus ini gulma yang digunakan adalah jawan (Echinochloa cruss-galli) dan kremah (Alternanthera sessilis)	Penggunaan metode ekstrasi ciri , dan metode klasifikasi yang lain untuk membandingkan tingkat *accuracy*.
7.	Perbandingan Metode PCA dengan HMM (Syakhala, Puspitaningrum, & Purwandari, 2015)	Hasil pengenalan wajah terbaik didapatkan dari citra wajah manusia tanpa background dengan tingkat *accuracy* tertinggi pada metode PCA sebesar 86,6% sedangkan *accuracy* metode HMM sebesar 77,7% selain itu PCA lebih cepat dalam pengenalan wajah dari 1 - 1,5 detik sedangkan pada metode HMM membutuhkan waktu 2 - 7,5 detik	Metode PCA dapat diimplementasikan dengan metode lain yang memiliki tingkat *accuracy* tinggi.
8.	Klasifikasi Hepatitis C (Suwardika, 2017)	Pada pengklasifikasian ini, metode CART , K-means, dan *Kernel* K-means menghasilkan ketepatan klasifikasi yang berbeda, dimana dari beberapa metode tersebut *kernel* K-means menunjukan bahwa	Gunakan metode klasifikasi lain dengan nilai *accuracy* jauh lebih tinggi seperti SVM.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		data hepatitis memiliki pengelompkan yang baik	
9.	Klasifikasi Hepatitis C (Soliman & Elhamd, 2014)	Studi komparasi dengan sistem lain menunjukan dimana PCA yang dikombinasikan dengan LS SVM memperoleh *accuracy* klasifikasi 98,86% sedangkan sistem lain memperoleh *accuracy* 96,12%	Penelitian kedepan dapat menggunakan teknik optimasi, lainnya. Selain itu, fungsi kernel lainnya dapat diterapkan pada fase klasifikasi.
10.	Klasifikasi Hepatitis C (A. Jilani, Yasin, & Mohammad Yasin, 2011)	Pada paper ini, dari 19 atribut, 6 tampaknya menjadi faktor yang paling penting. Pada sistem klasifikasi penelitian ini Keakuratan yang diperoleh dari proses yang kami terapkan adalah 100% untuk data pengujian dan 99,1% untuk data pelatihan.	Kedepannya faktor-faktor lain seperti enzim hati gambar darah, analisis urin (*bilirubinuria, proteinuria ringan, urobilinogen*), serologi untuk penanda virus dll harus dipertimbangkan. Untuk diterapkan *pada fuzzy learning model* pada data pasien.
11.	Oversampling SMOTE (Ramezankhani et al., 2016)	Untuk klasifikasi yang menggunakan *balanced dataset* (sudah dilakukan *oversampling*). menunjukkan bahwa klasifikasi yang dihasilkan sangat optimal dengan kenaikan nilai *sensitivity* dan *specificity* yang awalnya sama yaitu 20% persen menjadi 53% dan 73%	Kedepannya, implementasi metode SMOTE harus fokus terhadap *performance measures, optimal oversampling rate*, dan pemilihan klasifikasi terbaik untuk prediksi.

Dari Tabel 2.1 terdapat beberapa landasan masalah seperti klasifikasi Hepatitis C, perbandingan metode PCA dan ELM, metode SMOTE, dan kombinasi PCA dan ELM. Dari landasan terori tersebut dapat disimpulkan bahwa pengunaan metode klasifikasi ELM cocok untuk data medis, dan metode PCA dan ELM menghasilkan *accuracy* yang bagus ketika dikombinasikan dengan SMOTE. Sehingga pada penelitian ini menggunakan kombinasi metode SMOTE dan PCA pada algoritma ELM untuk melakukan klasifikasi penyakit Hepatitis C.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2 Virus Hepatitis C (HCV)

Hepatitis C disebabkan oleh infeksi virus Hepatitis C (HCV) dan tergolong penyakit yang menyerang hati. Hepatitis C menjadi penyebab utama sirosis dan penyakit hati kronis, serta inidikator umum penyebab transplantasi hati di banyak negara. Ada sekitar 170 juta orang yang terinfeksi di seluruh dunia, yaitu sekitar 3% dari populasi global (Brataatmadja, 2003). Virus ini berantai *single strand RNA* atau tunggal yang memiliki ukuran 50-60 *nm* dengan *genom positive-strand RNA* yang panjangnya 9400 *nukleotida*(*nt*). Secara *genotipe* ada 7 tipe virus hepatitis didunia. VHC I – VHC VII dimana tipe-tipe tersebut terbentuk sesuai dengan *region genomnya* berdasarkan susunan persamaan *nukleotida* (Puspaningrum, 2008). Virus hepatitis C termasuk virus paling berbahaya dari semua jenis virus hepatitis. Virus ini menyebabkan pembengkakan dan kemerahan didalam hati. Menurut pakar dibidang kesehatan, tingkat kematian akibat hepatitis meningkat tiga kali dalam dua puluh tahun ke depan (Soliman & Elhamd, 2014).

2.3 Infeksi HCV

HCV memiliki beberapa cara *transmisi* (penularan) seperti : kontak seksual, darah, *transmisi vertikal* contohnya ibu kandung yang didiagnosa HCV akan menurunkan ke anaknya, dan infeksi *nosocomial* yang diakibatkan kurangnya teknik sterilisasi pada alat medis (Cacoub & Comarmond, 2017). Faktor penyebab cepat berkembangnya hepatitis C dibedakan menjadi 3 yaitu lingkungan, virus, dan inang. Faktor lingkungan yang dipengaruhi tingginya konsumsi obat atau alkohol, *steatosis*. Faktor virus disebabkan berkembangnya *genotipe* dari HCV sedangkan faktor inang dipengaruhi oleh jenis kelamin, ras, umur, *genetic*, dan respon imun (Hairany, 2007).

2.4 Fibrosis Hati

Fibrosis hati terjadi karena akumulasi protein matriks ekstraselular (*Extracellular Matrix*, ECM). Hal itu merupakan ciri khas bagi kebanyakan jenis penyakit hati (kronik). Penyebab utama fibrosis hati seperti penyalahgunaan alkohol, HCV , dan *Non Alcoholic Steato Hepatitis* (NASH) (Bataller & Brenner, 2005).

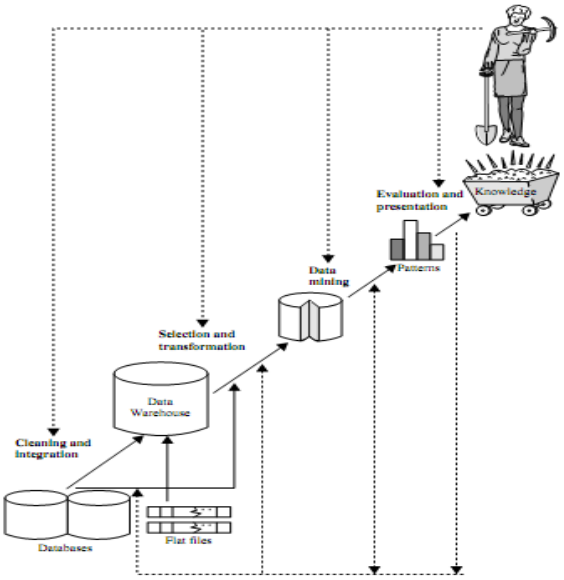
Sirosis merupakan tahapan akhir dari fibrosis hati yang menjadi masalah kesehatan terbesar di dunia. Ketika fibrosis berkembang menuju sirosis, ini akan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mengancam jiwa pasien yang menyebabkan komplikasi hati termasuk perdarahan varises, pembentukan asites, dan *sindrom hepatorenal* (Leija, Reyes, & Rodríguez, 2007). Menurut (Nasr, El-Bahnasy, Hamdy, & Kamal, 2018) fibrosis hati memiliki 5 tahapan yaitu *No Fibrosis*, *Portal Fibrosis*, *Few Septa*, *Many Septa*, dan *Cirrhosis*.

2.5 Data Mining

Data Mining merupakan sebuah teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data yang bertujuan untuk pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma. *Data mining* dapat digunakan untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya melalui proses analisis data dengan perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai. Definisi lain mengatakan *data mining* adalah pengolahan data berukuran besar yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan tiap tiap data. Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa *data mining* merupakan proses ekstraksi data menjadi informasi – informasi yang nantinya dapat digunakan yang berasal dari kumpulan data yang berukuran besar. *Data mining* dapat dibagi menjadi beberapa tahapan yang bersifat interaktif sebagai suatu rangkaian proses (Fridayanthie, 2015).



Gambar 2. 1 Tahapan *Data Mining*
(Mustafa, Ramadhan, & Thenata, 2018)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dari Gambar 2.1 dapat dijelaskan bahwa tahapan-tahapan *data mining* adalah sebagai berikut :

1. *Data Cleaning*. Proses *cleaning* pada data perlu dilakukan sebelum proses data mining dilaksanakan, hal ini dilakuakn untuk menentukan fokus *Knowledge Discovery in Database* (KDD). Proses *cleaning* mencakup beberapa hal seperti memeriksa data yang tidak konsisten, membuang duplikasi data, dan melakukan perbaikan pada kesalahan data, seperti kesalahan cetak.
2. *Data selection*. Proses seleksi dilakukan dari sekumpulan data operasional sebelum tahap penggalian. Data hasil seleksi akan disimpan dalam suatu berkas terpisah yang nantinya akan digunakan untuk proses *data mining*.
3. *Data Mining* merupakan proses implementasi teknik atau metode tertentu untuk mencari pola atau informasi menarik pada data. Pada *data mining* terdapat teknik, metode atau algoritma yang sangat bervariasi dan pemilihannya sangat bergantung pada tujuan dan jalannya proses secara keseluruhan.
4. *Evaluation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang dihasilkan dari proses *data mining* bertentangan dengan fakta, dan hipotesis sebelumnya atau tidak. Dan juga perlu ditampilkan dalam bentuk yang mudah dipahami oleh pihak terkait. .
5. *Knowledge* (pengetahuan). Merupakan tahapan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.6 Klasifikasi

Klasifikasi merupakan suatu proses analisis data yang mengekstrak model yang menggambarkan kelas data dan menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data satu dengan kelas data lainnya. Hal ini bertujuan untuk menyatakan objek tersebut masuk pada kategori yang sudah ditentukan sebelumnya (Iriadi, 2012).

Teknik klasifikasi bertujuan untuk menemukan fungsi keputusan $f(x)$ yang memiliki nilai keluar diskrit, dan secara akurat memprediksi kelas dari pola data yang dipelajari dengan pendekatan *supervised learning* untuk memprediksi data berikutnya yang memiliki kemiripan (Suwardika, 2017).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.7 K-fold cross validation

Cross-validation atau sering disebut estimasi rotasi merupakan sebuah evaluasi kinerja model atau algoritma menggunakan teknik validasi model dimana data dipisahkan menjadi dua bagian yaitu data proses pembelajaran (*training*) dan data proses pengujian (*testing*). Teknik ini bertujuan untuk memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dan utamanya digunakan untuk melakukan prediksi model dalam praktiknya. Salah satu teknik *cross-validation* ini adalah *k-fold cross validation*, dimana metode ini memecah *dataset* menjadi *k* bagian set data dengan ukuran yang sama. *k-fold cross validation* sering digunakan untuk menghilangkan bias pada data.

Gambar 2. 2 Model 3-fold cross validation
(Foeady, 2019)

Pada Gambar 2.2 merupakan penggunaan *3-fold cross validation*. Dimana data akan dibagi menjadi 3 subset data dan akan di eksekusi sebanyak 3 kali. Setiap subset data mempunyai kesempatan yang sama untuk menjadi data testing maupun data training. model Dengan diasumsikan nama setiap subset data yaitu K1, K2, dan K3 dapat dilakukan pengujian seperti berikut:

Pada percobaan pertama subset data K1 sebagai data testing (warna biru) sedangkan K2 dan K3 sebagai data training (warna putih).

Pada percobaan kedua subset data K2 sebagai data testing (warna biru) sedangkan data K1 dan K3 sebagai data training (warna putih)..

Pada percobaan ketiga atau terakhir subset data K3 sebagai data testing (warna biru) sedangkan K1 dan K2 sebagai data training (warna putih)..

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Performance klasifikasi (*accuracy*) dapat diukur dengan cara membandingkan seluruh data uji yang diklasifikasi benar dengan banyaknya data uji. Persamaan 1 adalah rumus yang digunakan untuk mengukur tingkat *accuracy* hasil klasifikasi (Dewi, 2018) .

$$accuracy = \frac{\sum klasifikasi\ benar}{\sum data\ uji}\ x\ 100\ \% \qquad (1)$$

Selanjutnya akan dilakukan perhitungan simpangan baku (*standar deviation*), simpangan baku adalah tingkat persebaran data yang menunjukkan jarak rata-rata dari nilai tengah (*median*) dengan suatu titik nilai pada data. Semakin besar hasil dari simpangan baku, maka tingkat persebaran dari nilai tengahnya juga besar, dan juga sebaliknya. Tujuan perhitungan simpangan baku pada proses ini adalah untuk mengetahui besarnya selisih antara ratarata *accuracy* dengan *accuracy* disetiap percobaan. Rumus untuk menghitung simpangan baku seperti pada Persamaan 2 (Brown, 1982).

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \qquad (2)$$

Keterangan :

σ = simpangan baku

N= banyaknya percobaan

x = percobaan ke-i

i = indeks setiap percobaan

μ = mean

2.8 Principal Component Analysis

Principal Component Analysis (PCA) merupakan sebuah metode untuk mengurangi beberapa fitur yang tidak diperlukan pada *dataset*. Biasanya struktur data akan direpresentasikan melalui grafik. Penggunaan metode ini akan sangat berguna sekali untuk data yang memiliki dua dimensi atau lebih (Santosa, 2007). Misalkan ada sebuat *dataset* dengan struktur sebagai berikut:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ x_{m1} & x_{m1} & \cdots & x_{mn} \end{bmatrix} \qquad (3)$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dimana n menjelaskan jumlah fitur/atribut dan m adalah banyaknya percobaan. Selanjutnya dengan cara *centering* yaitu mengurangi setiap data dengan rata-rata dari setiap atribut yang ada, data dapat ditransformasi ke dalam suatu kolom dengan mrnggunakan rumus sebagai berikut :

$$\hat{X} = X - \bar{X} \tag{4}$$

Dimana $\hat{X}$ merupakan vektor hasil dari proses *centering*, X adalah vektor kolom dan $\bar{X}$ adalah rata-rata dari jumlah kolom yang terkait. Proses tersebut dilakukan untuk semua kolom dari $i = 1$ sampai $i = n$. Setelah itu digunakan persamaan matriks *covariance* (C) untuk mendapatkan besaran baru. Persamaan tersebut dapat dilihat seperti berikut :

$$C = \frac{\hat{X}^T - X}{m-1} \tag{5}$$

Setelah mendapatkan nilai dari matriks *covariance*, langkah selanjutnya mencari nilai *eigenvalue* dan *eigenvector*. Dimana *eigenvalue* meupakan nilai yang didapat dengan mengurutkan mulai yang terbesar sampai yang terkecil, dan *eigenvector* diketahui dengan mengurutkan juga nilai yang bersesuaian dengan *eigenvalue* (Ismawan, 2015). Dari list tersebut kita akan mengetahui fitur yang memiliki nilai tinggi dan mana fitur yang memiliki nilai rendah. Dimana fitur yang memiliki nilai rendah bisa kita buang.

2.9 Extreme Learning Machine

Extreme Learning Machine (ELM) merupakan metode klasifikasi yang diterbitkan pada 2006, namun sudah diperkenalkan terlebih dulu pada tahun 2004 oleh Huang (Hidayat, 2012). Metode ini terdiri dari 3 *layer neural network* dengan menggunakan sebuah fungsi aktivasi pada proses perhitungannya. Dibandingkan dengan metode berbasis gradien konvensional, ELM memiliki banyak kelebihan seperti mampu meminimalisirkan jumlah iterasi, proses *learning* (pelatihan) yang sangat cepat, mampu megoptimalkan nilai *learning rate*, cocok untuk banyak fungsi aktivasi nonlinear dan memiliki nilai hasil klasifikasi yang lebih baik

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dibandingkan *backpropagation* dalam banyak kasus (Prakoso & Wisesty, 2016). Dalam prosesnya, ELM menggunakan jaringan *feed-fordward neural network* yang berbasis *single hidden layer* (SLFNs). Lapisan atau layer pada ELM terdiri dari *input layer, hidden layer, dan output layer* (Agustina, Anggraeni, & Mukhlason, 2005).

2.9.1 Feed Forward Algorithm

Algoritma *feedforward* memiliki proses yang hanya melalui propagaasi maju (*feed forward*) tanpa menggunakan propagasi mundur (*back forward*) merupakan algoritma yang memiliki beberapa *layer* (*multilayer*) yang terdiri dari beberapa *neuron* sebelum menuju perhitungan akhir (*output*). Algoritma *feedforward* bertujuan untuk mengenali pola suatu data sehingga dalam proses identfikasi atau klasifikasi memiliki Untuk mendapatkan hasil yang akurat algoritma *feedforward* memiliki tujuan mengenali pola suatu data dalam proses identfikasi atau klasifikasi (Huang, Zhu, & Siew, 2006). Dalam proses perhitungan, *feedforwad* akan menggunakan pola dari suatu fungsi aktifasi untuk mencari nilai bobot terbaik. Penerapkan fungsi aktifasi dan pejumlahan nilai bobot akan dilakukan disetiap lapisan atau *hidden layer*. Fungsi aktifasi yang paling umum digunakan adalah *sigmoid* karena dianggap mudah dan memberikan hasil baik. Nilai suatu bobot akan dioptimalkan pada proses ditiap-tiap *layer* sehingga hasil keluaran dari algoritma *feedforward* adalah nilai bobot paling optimal untuk mengidentifikasi beberapa kelas (Pandjaitan , 2007). Untuk mendapatkan nilai bobot paling optimal bobot pada tiap *layer* dapat digunakan Persamaan dibawah ini

$$H_{init} = X.W^t \tag{6}$$

Dengan :

H_{init}= Matriks inisialisasi *output hidden layer*

X = Matriks Hidden *Output*

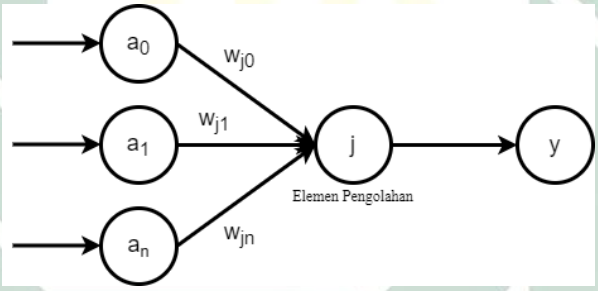
W^t= Matrik nilai Bobot

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Setelah mendapatkan nilai S_j, maka dilakukan perhitugan menggunakan fungsi *sigmoid* dengan persamaan *sigmoid* untuk mendapatkn nilai matrik H (*Hidden Output*) dengan persamaan berikut :

$$H = \frac{1}{1+e^{H_{init}}} \tag{7}$$

Dari perhitungan tersebut akan didapatkan nilai dari matrik *hidden output*. Algoritma *feedforward* memiliki sebuah arsitektur yang dapat dilihat melalui Gambar 2.2. Pada gambar tersebut terlihat bahwa *layer input* yang disimbolkan a_i akan melalui proses perhitungan penjumlahan bobot dan pembentukan fungsi pada *hidden layer* yang disimbolkan j dengan menggunakan fungsi aktifasi *sigmoid*. Setelah proses pada *layer j* selesai, akan dihasilkan suatu *output* yang disimbolkan y_i dimana berisi tentang informasi prediksi kelas data hasil klasifikasi (Huang et al., 2006).



Gambar 2. 3 Arsitektur Feedforward
(Foeady, 2019)

Semakin banyak jumlah *hidden layer* pada algotima *feedforward*, akan meningkatkan tingakat keberhasilan metode ini. Disisi lain jika terlalu banyak dapat menimbulkan permasalahan yakni pembaharuan bobot yang sudah optimal sehingga bobot tersebut Kembali menjadi tidak optimal (Huang et al., 2006).

2.9.2 Single Hidden Layer Feed Forward Neural Network (SLFNs)

SLFN memiliki fungsi *output* dengan jumlah sebanyak m dan fungsi aktivasi *input*. Sebuah matriks H (*hidden output*) dapat disusun jika jumlah *input* data sebanyak n dengan jumlah *hidden layer* yang diinisiasikan sebanyak m, dimana matriks berisi *output* dari hasil perhitungan pada *hidden layer* yang berukuran $m \times n$. Matriks tersebut dapat dilihatpada persamaan :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$H = \begin{bmatrix} g(x_1 \cdot w_1, 1 + b_1) & \cdots & g(x_1 \cdot w_1, m + b_m) \\ \vdots & \ddots & \vdots \\ g(x_n \cdot w_n, 1 + b_1) & \cdots & g(x_n \cdot w_n, m + b_m) \end{bmatrix} \quad (8)$$

Setelah matrik *hidden output* didapatkan selanjutnya akan mencari matriks *Moore–Penrose inverse* dengan persamaan sebagai berikut.

$$H^+ = (H^t . H)^{-1} . H^+ \quad (9)$$

Dimana :

H^+ = *Matriks Moore- Penrose inverse*

H = *Matriks Hidden Output*

H^t = Tranpose matrik $Hidden\ Output$.

Selanjutnya dari matrik H^+ dapat dicari nilai dari bobot *output* dengan rumus berikut :

$$\beta = H^t y \quad (10)$$

Dimana :

β = Nilai bobot *output*

H^+ .= Matriks *Moore- Penrose inverse*

y = Target atau *class label*

Proses terakhir dari training klasifikasi menggunakan metode ELM adalah menentukan nilai *output* dengan persamaan dibawah:

$$y = \beta H \quad (11)$$

Dimana :

y = Target atau *class label*

H = Matrik Hidden *Output*

β = Nilai bobot *output*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Sampai disini proses training ELM sudah selesai. Dan menghasilkan suatu model klasifikasi yang bisa digunakan untuk menguji suatu data untuk diklasifikasin atau biasa disebut proses testing. Proses testing akan menggunakan matrik X yang merupakan data testing. Lalu akan dilakukan perhitungan untuk mencari matrik *inisialisasi output hidden layer* dengan rumus berikut :

$$H_{init} = X.W^t \tag{12}$$

Dengan :

H_{init}= matrik *inisialisasi output hidden layer*

X = Matrik Hidden *Output*

W^t= Matrik nilai Bobot

Setelah itu mencari nilai matrik hidden *output* dengan aktivasi fungsi sigmoin seperti pada rumus berikut :

$$H = \frac{1}{1+e^{H_{init}}} \tag{13}$$

Prses terakhir adalah mencari *output* dari data *testing*. Karena ini ELM klasifikasi ilai y yang didapatkan harus dibulatkan. Rumus untuk mencari nilai *output* seperti berikut :

$$y = \beta H \tag{14}$$

2.9.3 Algoritma ELM

Huang mengatakan bahwa dalam metode ELM terdapat beberapa langkah-langkah yang harus diselesaikan, seperti menentukan *input* dari metode ELM itu sendiri, mempersiapkan data *training*, data *testing*, pemilihan fungsi aktifasi $g(x)$, dan menentukan jumlah *hidden layer* (Huang et al., 2006). Dimana proses tersebut dapat dijabarkan sebagai berikut:

1. Menentukan bias dan bobot *vector* secara acak sebagai faktor yang mempengaruhi nilai hidden layer.

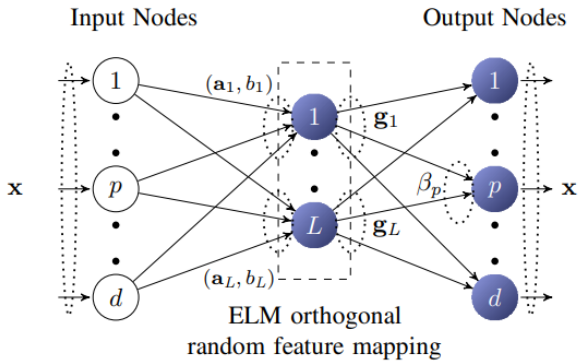
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Menggunakan Persamaan 12 unutuk enghitung atau mendefinisikan *output matriks* yang berupa $H_{n\times m}$ pada data training.
3. Menggunakan Persamaan 14 untuk menghitung *output* bobot β pada data *training*.
4. Dengan *output* nilai β yan gsudah didapatkan selanjunya dilakukan perhitungan *output* klasifikasi yang dihasilkan pada proses *training* dengan menggunakan data *testing*.

Penentuan bobot memang dilakukan secara *random*, Namun agar mendapatkan hasil yang lebih maksimal perlu adanya pembatasan nilai minimal dan maximal bobot. Umunya nilai bobot yang mendekeati 0 memberikan hasil klasifikasi yang lebih baik(Utomo & Kardiana, 2014).

2.9.4 Struktur Jaringan Syaraf Tiruan ELM

Secara umum jaringan syaraf tiruan ELM memiliki 3 *layer* yaitu *input layer, hidden layer, dan output layer*. dengan nilai *input* yang disimbolkan x , jumlahnya n dan jumlah *neuron* pada hidden *layer* disimbolkan m. Misalkan $x = [x1, x2, x3, \ldots, xn]$ dengan xi adalah inputan data untuk dilakukan proses ELM, selanjutnya membentuk matriks H dengan ukuran $n\ x\ m$ dengan melakukan proses penjumlahan bobot dan fungsi aktifasi $g(x)$. Lalu mencari matriks *output* T dengan dilakukan perhitungan yang sama dari hasil pada *hidden layer*. Karena nilai *output* hanya memiliki satu nodes saja ukuran matriks H adalah $m\ x\ 1$. Arsitektur ELM dapat dilihat pada Gambar 2.4.



Gambar 2. 4 Arsitektur *Extreme Learning Machine*
(Kasun, Zhou, Huang, & Vong, 2013)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.10 Python

Python adalah secara umum merupakan penulisan kode bahasa pemrograman yang memiliki tujuan umum (*general*) untuk dapat digunakan pada segala kebutuhan pemrograman. *Python* menjadi bahasa pemrograman populer yang digunakan di akademik maupun industri karena diciptakan oleh Guido van Rossum pada tahun 1990 di Belanda dengan kelebihan memiliki fitur *library* yang luas dan Bahasa yang mudah dan ringkas (Daniel, 2013).

Python merupakan salah satu bahasa pemrograman yang multiguna dan interpretatif, tidak seperti bahasa pemrograman lain yang sulit untuk dibaca dan dipahami. Hal ini dikarena *python* lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks sehingga sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain.

Python dikenal untuk kebutuhan sains dan teknologi. Salah satu penggunaannya adalah penerapan *boundary element method*. Library pendukung utama untuk kebutuhan tersebut adalah *NumPy*. Pada tingkat atas, *SciPy* sama dengan *NumPy*. Kedua *library* tersebut membantu dalam membuat dan manipulasi data. Kedua *library* ini membantu dalam mempercepat komputasi melalui optimalisasi *memory*. Sebagai contohnya, apabila pengguna ingin melakukan proses *looping* pada multidimensional *array* dalam jumlah besar maka proses *looping* akan berjalan cepat ketika pengguna menggunakan alokasi alamat memory pada masing-masing elemen *array* tadi. Mekanisme ini sudah ditangani oleh *Numpy* dan *SciPy*. Sebagai keuntungan tambahan, *library* ini sangat intuitif dan penjelasan yang mudah dipahami melalui manual atau dokumentasi

2.11 Confusion Matrix

Confusion matrix adalah sebuah teknik atau metode evaluasi sistem klasifikasi untuk mengetahui tingkat kesalahan klasifikasi pada data. Semakin kecil tingkat *error* yang dihasilkan, suatu sistem klasifikasi akan dinyatakan baik atau berhasil. Metode ini bertujuan mengetahui keakuratan dan tingkat *error* (kesalahan) dari suatu sistem klasifikasi (Visa, Ramsay, Ralescu, dan Van Der Knaap , 2011).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		Predicted				
		A_1	...	A_j	...	A_n
Actual	A_1	N_{11}		N_{1j}		N_{1n}
	⋮ A_i ⋮	N_{i1}	...	⋮ N_{ij} ⋮	...	N_{in}
	A_n	N_{n1}		N_{nj}		N_{nn}

Gambar 2. 5 *Confusion Matrix*
(Deng, Liu, Deng, & Mahadevan, 2016)

Gambar 2.5 memaparkan bentuk dasar dari *confusion matrix* yang digunakan untuk klasifikasi *multi-class*, dengan kelas A_1, A_2, A_3, dan A_n. Dalam *confusion matrix*, N_{ij} mewakili jumlah sampel yang sebenarnya dari kelas A_i, tetapi diklasifikasikan sebagai kelas A_j,

Pada *confusion matrix* nilai yang didapatkan dari *accuracy*, *sensitivity*, dan *specificity* menjadi pengukuran taraf evaluasi keberhasilan klasifikasi. Penjelasan dari tiap-tiap taraf evaluasi tersebut dapat dilihat dibawah ini:

1. *Accuracy*, merupakan nilai yang didapatkan dengan membandingkan banyaknya data yang terklasifikasi dengan benar dengan jumlah semua *dataset*, nilai tersebut menunjukkan keakuratan dari suatu klasifikasi. Semakin banyak data yang terklasifikasikan dengan benar membuktikan semakin baik nilai *accuracy*-nya. Rumus perhitungan *accuracy* dapat dilihat pada Persamaan :

$$Accuracy = \frac{\sum_{i=1}^{n} N_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} N_{ij}} \quad (15)$$

2. *Specificity* adalah suatu nilai yang merepresentasikan jumlah data yang benar, dan terklasifikasikan pada kelas negatif. Sebuah sistem klasifikasi dapat mengklasifikassikan kelas negatif dengan baik apabila nilai *specificity* semakin besar. Rumus perhitungan *specificity* dapat dilihat pada Persamaan :

$$\text{Specificity} = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ki}} \quad (16)$$

3. *Sensitivity* adalah suatu nilai yang merepresentasikan jumlah data yang benar dan terklasifikasikan pada kelas positif. Sebuah sistem klasifikasi dapat

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mengklasifikasikan data kelas positif dengan baik jika menghasilkan nilai sensitifitasnya semakin besar karena menunjukkan semakin besar pula data yang digunakan klasifikasi. Rumus perhitungan Sensitifitas dapat dilihat pada Persamaan :

$$\text{Sensitivity} = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ik}} \tag{17}$$

2.12 Synthetic Minority Oversampling Technique

Oversampling ini digunakan untuk menyelesaikan permasalahan *Imbalanced* data (ketidakseimbangan data), karena *dataset* dianggap tidak seimbang jika salah satu kelasnya memiliki dominasi yang sangat besar dibandingkan dengan kelas lainnya. Salah satu metode modifikasi dari *oversampling* adalah *Synthetic Minority Oversampling Technique* (SMOTE).

Metode ini diusulkan pertama kali pada tahun 2002 oleh Chawla, dimana kelas minoritas dilakukan *oversampling*-kan dengan membuat data *training* sintetis. Data *training* sintetis tersebut dibuat berdasarkan *k-nearest neighbor*. Pembangkitan data *training* sintetis yang berskala numerik berbeda dengan kategorik. Data numerik diukur berdasarkan jarak kedekatannya dengan jarak *Euclidean* sedangkan data kategorik dengan memperhitungkan nilai modusnya.

2.13 Exploratory Data Analysis

Exploratory Data Analysis (EDA) adalah bagian dari proses *data science*. EDA menjadi sangat penting sebelum melakukan *feature engineering* dan *modeling* karena dalam tahap ini kita harus memahami datanya terlebih dahulu. Metode ini memungkinkan *analyst* memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi dan lainnya.

2.14 Integrasi Keilmuan

Menjaga kesehatan agar terhindar dari berbagai penyakit sangat penting untuk dilakukan. Saat ini, banyak sekali penyakit yang berbahaya bagi kesehatan manusia. Salah satunya adalah penyakit fibrosis hati. Seperti namanya penyakit ini menyerang organ hati. Berdasarkan hasil wawancara yang dilakukan dengan pakar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

tafsir ibu Wiwin Luqna yang merupakan dosen Pendidikan Agama Islam UINSA setuju bahwa didalam islam proses klasifikasi hepatitis C yang bertujuan untuk memprediksi tingkat fibrosis hati berkaitan dengan menjaga kesehatan pada hati manusia. Banyak orang menganalisis bahwa orang yang sakit liver atau hepatitis itu disebabkan karena hatinya tidak tenang, banyak fikiran, ambisius, dan lain lain. Hati manusia harus dibersihkan dulu dari hal-hal yang merusaknya. dari segi kesehatan pun dijelaskan juga fungsi hati adalah membersihkan tubuh dari zat-zat yang berbahaya..

Di era modern seperti ini, metode *data mining* klasifikasi sering diterapkan pada data tertentu untuk mencari model sehingga bisa dilakukan prediksi. Dalam dunia kesehatan prediksi memiliki peranan yang sangat penting untuk mencegah penyakit agar tidak menjadi lebih parah (kronis). Seperti pada penelitian ini yang melakukan klasifikasi pada penyakit Hepatitis C untuk memprediksi tingkat fibrosis hati yang dialami. Kita sebagai manusia tentunya harus menjaga diri agar terhindar dari penyakit tersebut salah satunya melakukan pencegahan seperti tujuan utama penelitian ini.

Didalam al-qur'an juga sudah dijelaskan bahwa kita sebagai manusia harus dapat mencegah diri kita agar terhindar dari penyakit hati, maka hati harus terbebas dari hal-hal yang mengotorinya. Dari pandangan agama maka kita harus mengembalikan hati pada fungsinya. Fungsi hati menimbulkan daya rasa tenang, santun dan kasih sayang, lapang dada, taqwa, khusyu' dll. Didalam al-qur'an pun sudah dijelaskan berbagai cara agar terhindar dari penyakit hati seperti firman Allah sebagai berikut:

ۚ وَمَا جَعَلَهُ ٱللَّهُ إِلَّا بُشْرَىٰ وَلِتَطْمَئِنَّ بِهِۦ قُلُوبُكُمْ ۚ وَمَا ٱلنَّصْرُ إِلَّا مِنْ عِندِ ٱللَّهِ
إِنَّ ٱللَّهَ عَزِيزٌ حَكِيمٌ

Artinya : "*Dan Allah tidak menjadikannya (mengirim bala bantuan itu), melainkan sebagai kabar gembira dan agar hatimu menjadi tenteram karenanya. Dan kemenangan itu hanyalah dari sisi Allah. Sesungguhnya Allah Maha Perkasa lagi Maha Bijaksana*". *(Qs. Al Anfal: 10)*

Ayat diatas memberikan petunjuk bahwa hati kita harus selalu tenang tidak boleh gelisah dan cemas .Kehidupan seseorang pasti selalu berubah, Namun, ada satu hal yang tidak boleh pernah berubah pada diri kita, yaitu hati yang selalu

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

tenang, damai, dan tetap teguh dalam kebenaran. Selain itu kita juga harus selalu berlapang dada seperti firman Allah sebagai berikut:

ٱللَّهُ نَزَّلَ أَحْسَنَ ٱلْحَدِيثِ كِتَٰبًا مُّتَشَٰبِهًا مَّثَانِيَ تَقْشَعِرُّ مِنْهُ جُلُودُ ٱلَّذِينَ يَخْشَوْنَ رَبَّهُمْ ثُمَّ
تَلِينُ جُلُودُهُمْ وَقُلُوبُهُمْ إِلَىٰ ذِكْرِ ٱللَّهِ ۚ ذَٰلِكَ هُدَى ٱللَّهِ يَهْدِى بِهِۦ مَن يَشَآءُ ۚ
وَمَن يُضْلِلِ ٱللَّهُ فَمَا لَهُۥ مِنْ هَادٍ

Artinya : "*Allah telah menurunkan perkataan yang paling baik (yaitu) Al Quran yang serupa (mutu ayat-ayatnya) lagi berulang-ulang, gemetar karenanya kulit orang-orang yang takut kepada Tuhannya, kemudian menjadi tenang kulit dan hati mereka di waktu mengingat Allah. Itulah petunjuk Allah, dengan kitab itu Dia menunjuki siapa yang dikehendaki-Nya. Dan barangsiapa yang disesatkan Allah, niscaya tak ada baginya seorang pemimpinpun*." *(Qs. Az Zumar: 10)*

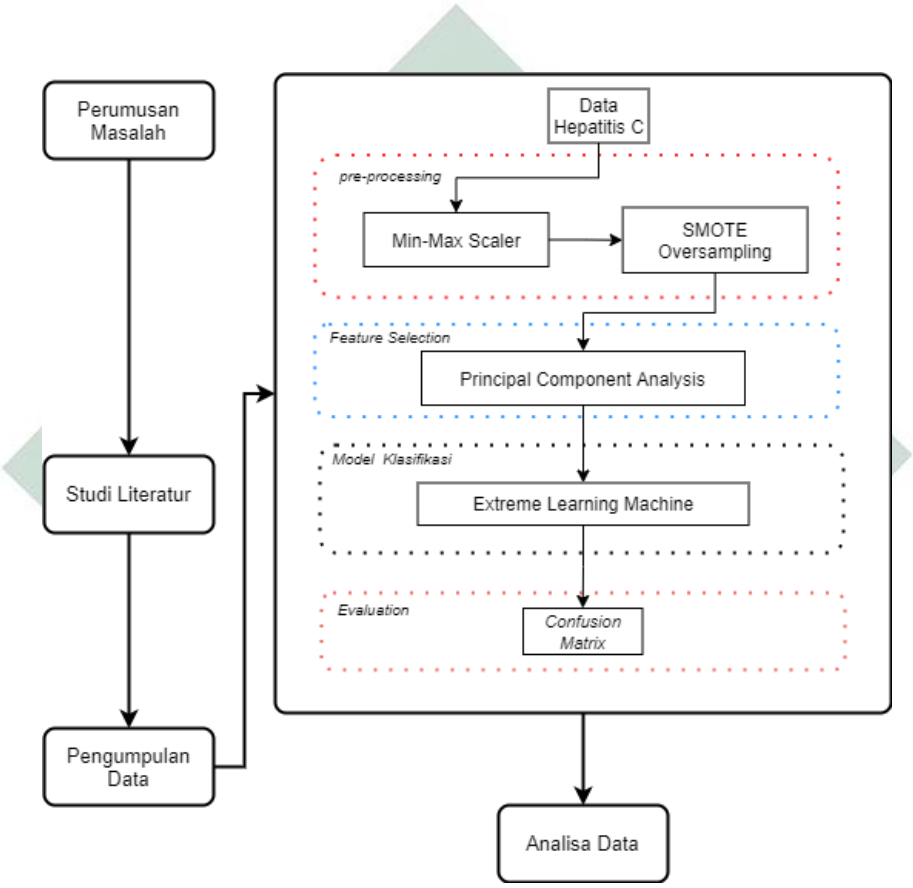
Dari penjelasan ayat diatas orang orang yang mau berlapang dada (sabar) pasti akan mendapatkan kenikmatandan terhindar dari berbagai jenis penyakit, termasuk penyakit hati. Dari paparan diatas dapat disimpulkan bahwa islam mengajarkan pada kita cara menjaga hati yaitu dengan selalu berlapang dada, harus memiliki fikiran yang tenang dan khusyu' dalam mengingat allah.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III
METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian akan dipaparkan dalam bentuk diagram alur, hal ini bertujuan unutk memudahkan pembaca dalam memahami rincian dari tiap langkah langkah penelitian. Metode penelitian yang digunakan terdapat pada Gambar 3.1.



Gambar 3. 1 Alur Penelitian

3.1.1 Perumusan Masalah

Perumusan masalah akan menjadi landasan dalam penelitian ini. Masalah yang di angkat pada penelitian ini seperti halnya tertera dalam latar belakang yakni mengenai klasifikasi fibrosis pada penderita hepatitis C. klasifikasi ini bertujuan untuk melakukan prediksi berdasarkan data pasien terdahulu. Dengan adanya klasifikasi ini pasien penderita hepatitis C dapat segera melakukan tindakan pencegahan sebelum memasuki fase yang lebih parah yatu gagal hati.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3.1.2 Studi Literatur

Tahapan ini dilakukan pembelajaran penelitian terdahulu terkait dengan klasifikasi penyakit hepatitis dan juga implementasi metode *Principal Component Analysis, Synthetic Minority Oversampling Technique,* dan *Extreme Learning Machine* pada klasifikasi *dataset* medis, Juga tentang HCV (Hepatitis C Virus) dan juga dampaknya, yaitu fibrosis hati. Sumber yang dapat digunakan sebagai studi literatur seperti jurnal, buku, dokumen, dan internet.

3.1.3 Pengumpulan Data

Pada penelitian menggunakan dua *dataset* yang berasal dari sebauh situs penyedia data UCI repository yaitu *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang berasal dari Mesir (Learning, 2019) dengan jumlah data sekunder yang digunakan sebanyak 1385 record yang terdiri dari 28 atribut dan satu *class label*. Berikut atribut-atribut data yang digunakan yaitu:

1. Umur (*Age*), keberadaan suatu benda atau makhluk, baik yang hidup maupun yang mati yang dihitung menggunakan ukuran rentang waktu. Rentang yang digunakan yaitu 0 - 32, 32 - 37, 37 – 42, 42 – 47, 47 – 52, 52 – 57, 57 – 62.
2. Jenis-kelamin (*Gender*), Suatu pembedaan tingkat maskulinitas dan femininitas dari serangkaian karakteristik yang terikat pada hal tertentu. Karakeristik tersebut dapat mencakup jenis kelamin (*male or female*).
3. Indeks Masa Tubuh (*Body Mass Index*) adalah sebuah matriks standar dimana nilai angka digunakan untuk menentukan tingkat berat badan apakah ideal, kurang, atau berlebih. Dan sering digunakan unutk menentukan golongan orang yang memiliki berat badan sehat dan tidak sehat. Rentang yang digunakan adalah [0 - 18:5] [18:5 - 25], [25 - 30], [30 - 35], [35 - 40]
4. Demam (*Fever*), Hal ini biasa terjadi karena sistem kekebalan tubuh sedang melawan suatu virus atau bakteri yang dapat menimbulkan penyakit.
5. Mual/muntah (*Nausea/Vomiting*), *Vomitus* adalah istilah kesehatan untuk muntah. Saat sering mengalami mual/muntah harus segera diobati karena dapat berakibat dehidrasi. Saat mual/muntah isi perut akan terdorong keluar dari mulut akibat otot perut yang mengalami kontraksi.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

6. Sakit kepala (*Headache*) adalah rasa sakit yang diiringi nyeri yang muncul pada bagian kepala, sakit ini bisa muncul secara bertahap maupun mendadak. Dan juga sakit kepala yang sering dialami oleh seseorang bisa muncul disalah satu sisi maupun seluruh bagian kepala.
7. Diare (*Diarrhea*) sebuah penyakit yang biasanya terjadi paling sedikit tiga kali dalam 24 jam dan ditandai saat tinja atau feses yang keluar memiliki karakteristik cair dan lembek..
8. Kelelahan (*Fatigue*) merupakan suatu kondisi yang disertai dengan perasaan lemah dan letih karena berkurangnya kapasitas yang dimiliki seseorang untuk bekerja..
9. Nyeri tulang (*Bone ache*), suatu kondisi nyeri atau sakit yang terjadi pada tulang. Kondisi ini dapat terjadi pada satu atau lebih tulang anda dan sangat mempengaruhi tubuh.
10. Penyakit kuning (*Jaundice*), kondisi dimana kulit berubah menjadi kuning. Tak hanya pada kulit, warna putih pada mata juga dapat berubah menjadi kuning. Dalam kasusnya semakin parah penyakit ini ketika warna kuning menjadi coklat atau orange..
11. Nyeri epigastric (*Epigastric pain*) bukanlah diagnosis atas suatu penyakit melainkan gejala suatu penyakit yang berupa rasa nyeri di area perut bagian atas (ulu hati).
12. Sel darah putih (*White Blood Cells*) termasuk sebagai bagian dari sistem kekebalan tubuh berfungsi untuk membantu tubuh melawan berbagai penyakit infeksi. sel ini berbentuk komponen darah dan memiliki jumlah sekitar 7000-25000 sel per tetes pada manusia dewasa yang sehat. Namun jumlahnya dapat meningkat hingga 50000 sel per tetes pada manusia dengan kasus penyakit leukimia.
13. Sel darah merah (*Red Blood Cells*) berfungsi sebagai pengikat oksigen yang diperlukan untuk oksidasi jaringan-jaringan tubuh melalui darah. Merupakan jenis sel darah yang paling banyak dimiliki oleh manusia dan shewan bertulang belakang.
14. Hemoglobin memiliki tugas berfungsi sebagai pengangkut oksigen dari paru-paru ke seluruh tubuh.dan dalam istilah medis disebut *metaloprotein* (protein

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang mengandung zat besi) di dalam sel darah merah yang ada pada manusia, hewan mamalia dan hewan lainnya.

15. Keping darah, trombosit (*Platelet*), merupakan sel yang dapat menyebabkan pembekuan darah (trombus). Sel ini sering terlibat dalam mekanisme hemostatis dan berbentuk anuklear kecil yang beredar dalam darah. Jika jumlah sel ini tinggi dapat meningkatkan risiko trombosis dan disfungsi keping darah yang menyebabkan pendarahan.
16. *Transaminase aspartatat* (AST1)(week1). Pada tubuh manusia AST memiliki dua bentuk (isoenzim) yaitu AST *mitokondria* yang terdapat didalam tubuh dan AST dalm bentuk bebas. Selain pada organ hati, AST juga ditemukan di ginjal, otot rangka, dan jantung..

17-22. *Transaminase alanina* (ALT) sering dikaitkan dengan kinerja organ hati. Enzim ini dapat dijumpai di dalam serum darah dan berbagai jaringan tubuh. Ada sebuah kondisi serius yang merupakan gejala gangguan hati yaitu ketika terjadi peningkatan rasio serum ALT dan AST dalam rentang antara batas atas normal dan lima kali nilai batas. Kondisi tersebut dapat disebabkan oleh alkohol, toksin, obat-obatan. Hal terburuk yang dapat terjadi seperti infeksi viral akut (kronis), dan sirosis hati. Data ALT yang digunakan dimulai dari week1, week4, week12, week24, week36, hingga week48

23-27. Asam ribonukleat (RNA) molekul polimer yang terlibat dalam berbagai peran biologis dalam mengkode, dekode, regulasi, dan ekspresi gen. RNA dan DNA merupakan empat makromolekul utama yang penting untuk semua bentuk kehidupan yang diketahui. Pengmabilan smapel dilakukan pertama kali saat proses perawatan pasien. Data RNA yang digunakan mulai dari RNA base (awal perawatan), RNA week 4 (minggu ke-empat), RNA week 12 (minggu ke-empat), RNA EOT (akhir perawatan), dan RNA EF (Elongation Factor) proses dimana RNA akan membuka ikatan double helix pada bagian gen sehingga akan membentuk RNA baru. Enzim RNA akan terus membentuk mRNA hingga terbentuk rantai mRNA, Rantai mRNA ini akan terus memanjang. Oleh kare itu tahap ini disebut tahapan Elongation (pemanjangan)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

28. *Baseline Histological Grading*, merupakan sebuah pengukuran atau penilaian terhadap tingkat penyebaran penyakit fibrosis itu sendiri. Rentang grade dimulai dari 1 sampai dengan 16.
29. *Class label*, variable yang menunjukan tingkat fibrosis pasien penderita pasien hepatitis C, portal fibrosis, few septa, many septa, atau Sirosis.

Dataset kedua yaitu *HCV data Data Set* yang berasal Jerman (Learning, 2020) dengan jumlah data sekunder yang digunakan sebanyak 615 record yang terdiri dari 12 atribut dan satu *class label*. Atribut-atribut data yang digunakan yaitu:

1. Umur (*Age*), dengan rentang 19 tahun – 74 tahun
2. Jenis-kelamin (*Gender*) yang terdiri dari dua jenis yaitu *male* dan *female*
3. *Albumin* (ALB) adalah protein utama yang terdapat dalam darah manusia yang diproduksi oleh organ hati. Albumin berfungsi untuk mengatur tekanan dalam pembuluh darah dan menjaga agar cairan yang terdapat dalam pembuluh darah tidak bocor ke jaringan tubuh sekitarnya.
4. *Bilirubin* (BIL) adalah zat yang terbentuk secara normal dari proses penguraian sel darah merah di dalam tubuh. Zat inilah yang memberikan warna kuning pada tinja dan urine.
5. *Choline esterase* (CHE) adalah kelompok enzim yang dipecah menjadi *choline* dan *thiocholine esters*. Serum *Cholinesterase* dan *Pseudocholinesterase* juga umum digunakan sebagai nama pemeriksaan.
6. *Gamma Glutamyl Transferase* (GGT) merupakan enzim yang ditemukan di berbagai organ tubuh. Akan tetapi, kadar atau konsentrasi enzim ini paling tinggi ditemukan di organ hati.
7. *Transaminase Aspartatat* (AST). Pada tubuh manusia AST memiliki dua bentuk (*isoenzim*) yaitu AST *mitokondria* yang terdapat didalam tubuh dan AST dalm bentuk bebas.
8. *Transaminase Alanina* (ALT) sering dikaitkan dengan kinerja organ hati. Enzim ini dapat dijumpai di dalam serum darah dan berbagai jaringan tubuh.
9. *Alkaline Phosphatase* (ALP) merupakan enzim yang berfungsi untuk membantu memecah protein dalam tubuh.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

10. *Cholesterol* (CHOL) adalah lemak yang berguna bagi tubuh. Namun bila kadarnya di dalam tubuh terlalu tinggi, kolesterol akan menumpuk di pembuluh darah dan mengganggu aliran darah
11. *Creatinine* (CREA) dikeluarkan dari tubuh oleh ginjal, yang menyaring hampir semuanya dari darah dan melepaskannya ke urin.
12. *Protein* (PROT) adalah zat yang terbentuk secara alami dan sangat kompleks yang terdiri dari residu asam amino yang bergabung dengan ikatan peptida.

Dari dua *dataset* yang digunakan memiliki beberapa atribut yang serupa sehingga perbandingan dari kedua *dataset* tersebut dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Perbandingan Dataset

Nama Atribut	Dataset yaitu *Hepatitis C Virus (HCV) for Egyptian patients Data Set*	Dataset *HCV data Data Set*
Age	✓	✓
Gender	✓	✓
Body Mass Index	✓	
Fever	✓	
Nausea/Vomiting	✓	
Headache	✓	
Diarrhea	✓	
Fatigue	✓	
Bone ache	✓	
Jaundice	✓	
Epigastric pain	✓	
White Blood Cells	✓	
Red Blood Cells	✓	
Hemoglobin	✓	
Platelet	✓	
Transaminase aspartatat (AST)	✓	✓
Transaminase alanina (ALT)	✓	✓
Asam ribonukleat (RNA)	✓	
Baseline Histological Grading	✓	
Albumin (ALB)		✓
Bilirubin (BIL)		✓
Choline Esterase (CHE)		✓
Gamma Glutamyl Transferase (GGT)		✓

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

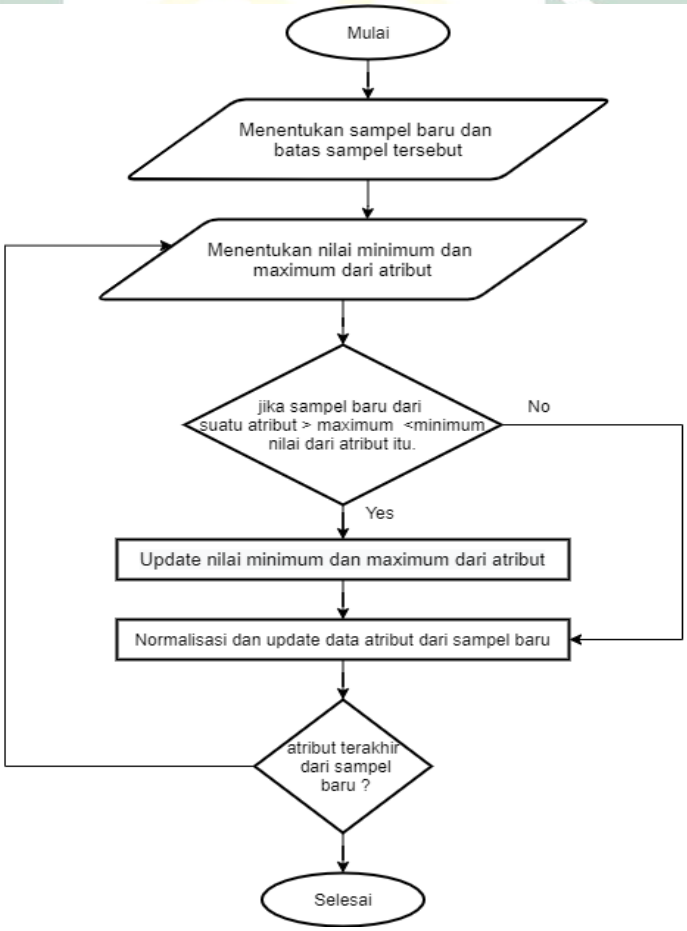
Alkaline Phosphatase (ALP)		✓
Cholesterol (CHOL)		✓
Creatinine (CREA)		✓
Protein (PROT)		✓

3.1.4 Pengolahan Data

Setelah data diperoleh, langkah selanjutnya adalah tahap pengolahan data. Pada penelitian ini pengolahan data terbagi menjadi 3 tahapan, *pre-processing*, klasifikasi, *dan evaluating*.

a. Pre Processing

Pada *pre-processing* data akan melalui dua tahapan proses, pertama dilakukan *scaling* menggunakan algoritma *min max scaler*. Tahapan ini bertujuan untuk mengatasi permasalah *interval dataset* yang cukup jauh. *Interval* jauh tersebut akan menyulitkan algoritma ELM untuk melakukan proses *train*. *Flow* dari algoritma *min-max scaler* dapat dilihat pada Gambar 3.2.



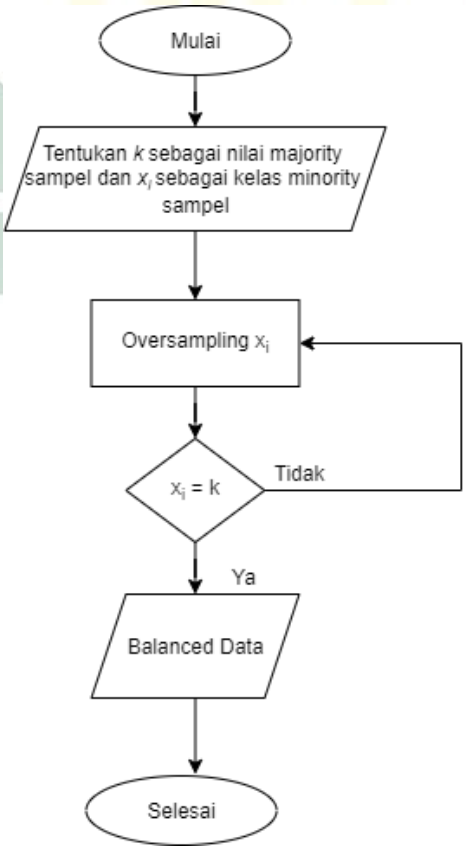
Gambar 3. 2 *Flow Min-max Scaler*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Berdasarkan Gambar 3.2 tahapan *scaling* algoritma *min-max scaler* adalah sebagai berikut :

3. Menetukan jumlah sampel baru dan juga nilai batas dari sampel tersebut
4. Tentukan nilai maksimum dan minimum pada atribut
5. Normalisasi dan update nilai atribut dari sampel jika nilainya lebih besar dari nilai maksimum dan lebih kecil dari nilai minimum
6. update nilai atribut dari sampel jika nilainya tidak lebih besar dari nilai maksimum dan tidak lebih kecil dari nilai minimum
7. Ulangi proses hingga mencapai sampel terakhir dari atribut

Tahapan proses kedua adalah *oversampling* dengan metode SMOTE, dimana data akan diperbanyak dengan cara membuat sebuah sampel baru yang memiliki karakteristik sama dengan sampel yang sudah ada dengan tujuan menyeimbangkan jumlah data tiap *class label*. *Flow* dari algoritma SMOTE dapat dilihat pada Gambar 3.3.



Gambar 3. 3 *Flow* SMOTE *Oversampling*

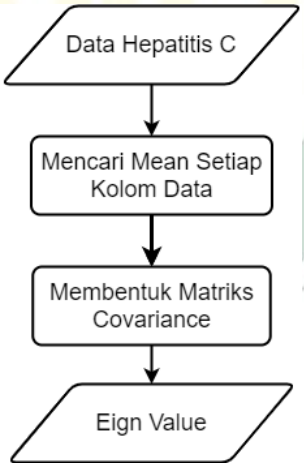
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Berdasarkan Gambar 3.3 tahapan *oversampling* algoritma SMOTE adalah sebagai berikut :

1. Tentukan *k* sebagai nilai *majority* dari suatu *class* dan x_i sebagai kelas *minority* sampel
2. *Oversampling* x_i
3. Ulangi proses tersebut hingga $x_i = k$

b. Featured Selection

Tahapan selanjutnaya adalah fitur seleksi yang merupakan proses memilih fitur-fitur yang relevan yang mempengaruhi hasil klasifikasi baik dari segi efektifitas maupun efisiensi kerja algoritma. Metode fitur seleksi yang akan digunakan adalah *principal component analysis* yang memiliki *flow* seperti Gambar 3.4.



Gambar 3. 4 *Flow* Metode *Principal Component Analysis*

Berdasarkan Gambar 3.4 tahapan seleksi fitur yang dilakukan adalah sebagai berikut :

1. Data akan akan melalui proses seleksi fitur menggunakan metode PCA
2. Mencari nilai rata – rata dari setiap kolom data dengan menggunakan Persamaan 4.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3. Menghitung nilai matriks covarian setelah rata – rata nilai setiap kolom didapatkan dengan persamaan 5.
4. *Eigen Value*, Tahapan terakhir ini untuk menentukan nilai akhir dari proses seleksi fitur PCA, yaitu dengan mencari nilai *eigen value*. Nilai *eigen value* akan diurutkan dari yang terkecil sampai yang paling besar. Selanjutnya ELM akan melakukan proses klasifikasi menggunakan nilai dari *eigen value* tersebut,

c. Proses Klasifikasi

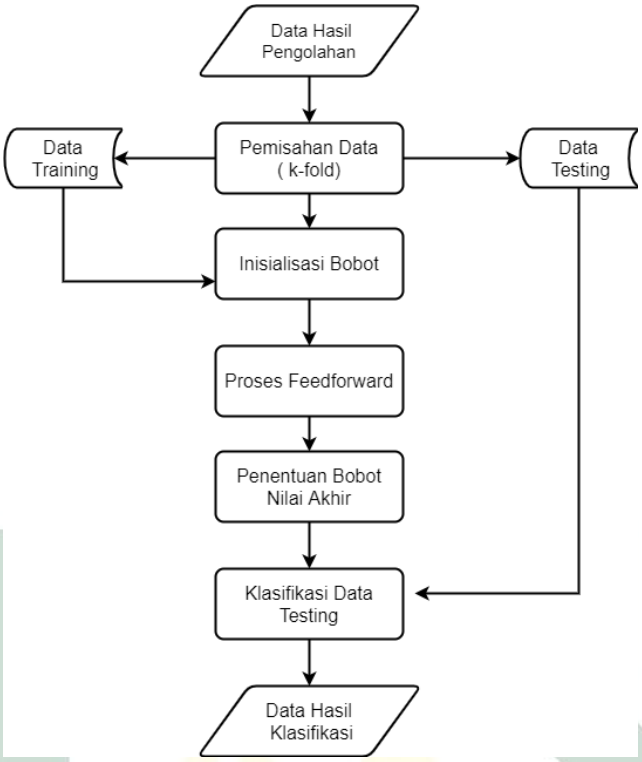
Pada tahapan ini akan menentukan tingkat keberhasilan penelitian ini. Tujuan dari proses klasifikasi akan melakukan perhitungan *dataset* yang sudah dilakukan *pre-processing* dan juga *featured selection* dengan menggunakan algoritma *extreme learning machine.* Untuk mendapatkan informasi penuh dari penerapan *pre-processing* dan juga *featured selection* pada proses klasifikasi ini akan dibagi menjadi 4 sekenario yang dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Skenario Klasifikasi ELM

ELM	PCA+ELM	SMOTE+ELM	SMOTE + PCA+ELM
Pada sekenario ini *dataset* akan diklasifikasikan dengan metode ELM tanpa dilakukan *oversampling* dan *featured selection.*	Fitur dari *dataset* akan direduksi dengan metode PCA yang selanjutnya akan diklasifikasikan dengan metode ELM	*Dataset* akan dilakukan *oversampling* dengan SMOTE sehingga jumlah *class data* akan seimbang lalu selanjutkan akan diklasifikasikan dengan ELM	Sekenario utama pada penelitian ini adalah menggabungkan metode SMOTE dan PCA untuk mendapatkan pengaruh hasil klasifikasi yang maksimal sesuai dengan metodologi penelitian

Dari beberapa sekenario pada Tabel 3.2 akan menghasilkan hasil klasifikasi yang nantinya bisa dibandingkan untuk ditarik sebuah kesimpulan dari tujuan utama penelitian ini. Secara umum *flow* proses klasifikasi metode ELM dapat dilihat pada Gambar 3.5.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 3. 5 *Flow* Metode *Extreme Learning Machine*

1. Pada proses *input dataset* yang digunakan adalah hasil dari seleksi fitur metode *Principal Component Analysis*

2. Pemisahan data bertujuan untuk membagi data menjadi dua kategori yaitu *training*, dan *testing* menggunakan metode *10-fold*.

3. Inisialisasi bobot dilakukan dengan menggunakan bilangan kecil dari -0,5 – 0,5 dan perhitungan akan dilakukan menggunakan persamaan 6. Setelah mendapatkan hasinya selanjutnya akan dilakukan perhitungan dengan fungsi aktivasi *sigmoid* dengan rumus pada persamaan 7.

4. Dengan menggunakan fungsi aktivasi, pada *feedforword* setiap unit *input* $(x_1, \ldots, x_n)$ menerima sinyal *input* lalu meneruskan sinyal tersebut pada lapisan tersembunyi *(hidden node)* yang biasa disebut matriks *Moore- Penrose inverse*

5. Dengan menggunakan persamaan 1 dapat dihitung bobot akhir dari hidden *layer* ke-*output layer* . Dimana H^+ merupakan matriks *Moore-Penrose Generalized Invers* dari matriks *H*, sedangkan matriks yang tersusun dari *output* masing-masing *Hidden Layer* merupakan matriks *H*, dan *T* adalah

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

matriks target. Selanjutnya menggunakan persamaan 11. untuk mencari nilai *output*.

6. Proses terakhir pada ELM adalah proses pengujian atau *testing*. Dimana model yang diperoleh dari proses *training* akan lakukan perhitungan dengan urut persamaan 12, persamaan 13, dan persamaan 14 dengan data testing.

d. Evaluating

Metode yang digunakan pada proses ini *confusion matrix*, dengan dilakukan 4 skenario untuk kebutuhan pengujian algoritma *Extreme Learning Machine*. Skenario pertama akan dilakukan klasifikasi metode ELM tanpa *oversampling* dan *featured selection*, kedua akan dilakukan evaluasi terhadap klasifikasi metode ELM dengan *oversampling* SMOTE, skenario ketiga akan dilakukan evaluasi terhadap klasifiaksi metode ELM dengan *featured selection* PCA, dan yang terakhir evaluasi klasifikasi metode ELM dengan kombinasi metode *oversampling* SMOTE dan *featured selection* PCA. Untuk pembagian data *traning* dan data *testing* akan dilakukan dengan metode *10-fold* agar hasil klasifikasi teruji validasi dan diyakini kebenarannya. Implementasi algoritma *10-fold* akan membagi data *traning* sebesar 90 % dan data *testing* sebesar 10%. Seperti namnya *10-fold* proses klasifikasi akan dilakukan sebanyak 10 kali dengan data *training* dan *testing* yang berbeda beda. Hasil tersebut akan dicari rata rata-ratanya dengan menggunakan Persamaan 1. yang akan menjadi final *accuracy* dari proses klasifikasi. Dengan metode *confusion matrix* dapat mengukur *specificity*, *accuracy*, dan juga *sensitivity* dengan menggunakan Persamaan 15, Persamaan 16, dan Persamaan 17.

3.1.5 Analisa Data

Tahapan ini akan membandingkan hasil klasifikasi dari dua *dataset* berbeda yaitu *Hepatitis C Virus (HCV) for Egyptian patients Data* Set dan *HCV data Data Set*. Meskipun ada sebagian atribut yang sama dari kedua *dataset* namun struktur datanya sangat berbeda. Dari perbandingan tersebut akan dapat diketahui diantara metode SMOTE, PCA, dan SMOTE+PCA yang memberikan pengaruh dalam peningkatan *specificity*, *accuracy*, dan juga *sensitivity* terbaik dengan perhitungan pada Persamaan 15, Persamaan 16, dan Persamaan 17 dalam algoritma *confusion matrix* dari serangkaian test yang telah diakukan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB IV
HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan hasil penelitian berdasarkan tahapan yang telah dijelaskan di metode penelitian pada bab sebelumnya. Tahapan tersebut mulai dari proses *pre-processing* atau pengolahan data, *training* model klasifikasi hingga evaluating dengan beberapa skenario yang telah ditentukan. Dan juga akan menjelaskan rancangan dan alur sistem klasifikasi yang telah dibuat beserta uji sistem yang dilakukan.

4.1. Hasil Klasifikasi ELM

4.1.1 Hepatitis C Virus (HCV) for Egyptian patients Data Set

4.1.1.1 Pre-Processing

Data *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang digunakan berjumlah 1385 data. *Dataset* tersebut harus bersih dari *noise* dan *valid* sebelum dilakukan proses klasifikasi dengan beberapa skenario yang telah disiapkan. *Dataset* tersebut haruslah sesuai dengan rancangan dan kebutuhan dari algoritma ELM dan bebsas dari permasalahan *dataset* seperti interval data. Dari kebutuhan tersebut maka diperlukan tahapan scaling pada proses ini

Dengan interval yang cukup jauh pada tiap atribut dari *dataset* pasien Mesir perlu dilakukan sebuah normalisasi dengan menggunakan algoritma *min max Scaler*. Hasil pengolahan yang telah dilakukan dapat dilihat pada Tabel 4.1.

Tabel 4. 1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set* Setelah *Pre-processing* dengan Min Max Scaler

Age	Gender	BMI	⋯	RNA EOT	RNA EF	Baseline histological Grading
82758621	0	1	⋯	0	0	76923077
48275862	0	53846154	⋯	41660101	03835484	07692308
86206897	0	84615385	⋯	91031548	68962692	07692308
⋮	⋮	⋮	⋰	⋮	⋮	⋮
34482759	0	30769231	⋯	42515446	19800871	23076923
68965517	0	53846154	⋯	00303669	8589966	92307692
79310345	1	30769231	⋯	0	0	76923077

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dataset telah dikonversi menjadi kisaran 0-ke-1, dan dapat diketahui bahwa nilai terbesar untuk setiap kolom sekarang memiliki nilai 92 dan nilai terkecilnya adalah

4.1.1.2 Klasifikasi

Dataset yang sudah melalui proses *pre-processing* dan juga *featured selection* akan diklasifikasikan dengan metode *extreme learning machine*. Aktivasi fungsi yang digunakan adalah *sigmoid*, dimana *sigmoid* ini akan menerima angka tunggal dan mengubah nilai *x* menjadi sebuah nilai yang memiliki *range* mulai dari 0 sampai 1. Algoritma ELM dengan aktivasi fungsi *sigmoid* akan dijelaskan dengan *pseudocode* berikut ini.

```
program ELM_ Hepatitis C Virus (HCV) for Egyptian patients Data
Set

1. import numpy
2. Tentukan jumlah neuron h pada hidden layer
3. Buat matriks nilai bobot awal W berukuran jumlah hidden
neuron x jumlah atribut
4. isi matriks nilai bobot dengan angka acak antara (-1 sampai
dengan 1 )
5. Hitung matriks inisialisasi output hidden layer dengan rumus
Hinit = X @ W.T
6. hitung outout dari proses aktivasi menggunakan sigmoid H=
1/(1 + np.exp(-Hinit))
7.  hitung Hplus= np.linalg.inv(H.T @ H) @ H.T
8. Hitung output weight dengan rumus beta = Hplus @ t
9. untuk mendapatkan output hasil y = H @ beta
```

Pada proses klasifikasi dengan metode ELM terdiri dari dua tahapan yaitu *training* dan *testing*. Dimana training dilakukan untuk membuat sebuah model klasifikasi sedangkan *testing* dilakukan untuk menentukan *output* dari hasil klasifikasi.

4.1.1.3 Evaluating

Hasil dari proses *learning* akan dilakukan proses evaluasi dengan menggunakan metode *confusion matrix*. Penggunaan *confusion matri*x berguna untuk menentukan keberhasilan suatu sistem klasifikasi. *Confusion matrix* sendiri memiliki 3 parameter keluaran yaitu *accuracy*, *sensitivity*, dan *specificity*. Semakin tinggi nilai yang dihasilkan maka akan semakin tinggi pula tingkat keberhasilkan dari sistem tersebut.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Evaluasi akan dilakukan pada sekenario klasifikasi dengan jumlah hidden *nodes* yang berbeda beda di tiap trainingnya guna mendapatkan hasil klasifikasi terbaik. Hasil dari evaluasi dari sekenario ini dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Hasil klasifikasi ELM *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	25%	26%	26%	26%	25%	25%	25%	26%	26%
	Sensitivity	25%	26%	26%	26%	25%	25%	25%	26%	26%
	Specificity	25%	26%	26%	26%	25%	25%	25%	26%	26%

Dari hasil klasifikasi ELM Tabel 4.2 hasil terbaik didapatkan dengan jumlah *accuracy* 26% dengan hidden nodes 100, 250, 300, 1000, dan 150 Dengan nilai nilai *sensitivity* dan *specificity* yang sama yaitu 26%.

4.1.2 HCV data Data Set

4.1.2.1 Pre-Processing

HCV data Data Set memiliki data berjumlah 615. *Dataset* tersebut juga dilakukan *pre-processing* agar sesuai dengan rancangan dan kebutuhan dari algoritma ELM dan bebsas dari permasalahan *dataset* seperti *interval data*. Oleh karena itu tahapan scaling akan dilakukan pada *HCV data Data Set*.

Scaling pada *HCV data Data Set* juga menggunakan algoritma *min-max scaler* karena *interval* data yang cukup berbeda meskipun tidak sebesar interval data *Hepatitis C Virus (HCV) for Egyptian patients Data Set*. Hasil implementasi dari algoritma *min max scaler* dapat dilihat pada Tabel 4.3.

Tabel 4. 3 *HCV data Data Set* setelah *Pre-processing* dengan *Min Max Scaler*

Age	Gender	ALB	⋯	CREA	GGT	PROT
22413793	0	35066865	⋯	09149473	0	76923077
22413793	0	35066865	⋯	0616189	03835484	07692308
22413793	0	47548291	⋯	07282233	68962692	07692308
⋮	⋮	⋮	⋰	⋮	⋮	⋮
77586207	1	20950966	⋯	05480347	09235767	82300885
46551724	1	26894502	⋯	04107926	07038985	57964602
68965517	1	31352155	⋯	05508356	04563738	51327434

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dataset telah dikonversi menjadi kisaran 0-ke-1, dan dapat diketahui bahwa nilai terbesar untuk setiap kolom sekarang memiliki nilai 92 dan nilai terkecilnya adalah

4.1.2.2 Klasfikasi

HCV data Data Set yang sudah melalui proses *pre-processing* akan diklasifikasikan dengan metode *extreme learning machine*. Aktivasi fungsi yang digunakan juga *sigmoid*, tahapan klasifikasi *HCV data Data Set* juga akan melalui beberapa sekenario untuk mendapatkan kesimpulan terbaik dari proses klasifikasi. algoritma ELM *HCV data Data Set* terlihat seperti dibawah ini.

```
program ELM_ HCV data Data Set

1. import numpy
2. Tentukan jumlah neuron h pada hidden layer
3. Buat matriks nilai bobot awal W berukuran jumlah hidden
neuron x jumlah atribut
4. isi matriks nilai bobot dengan angka acak antara (-1 sampai
dengan 1 )
5. Hitung matriks inisialisasi output hidden layer dengan rumus
Hinit = X @ W.T
6. hitung outout dari proses aktivasi menggunakan sigmoid H=
1/(1 + np.exp(-Hinit))
7.  hitung Hplus= np.linalg.inv(H.T @ H) @ H.T
8. Hitung output weight dengan rumus beta = Hplus @ t
9. untuk mendapatkan output hasil y = H @ beta
```

Pada proses klasifikasi dengan metode ELM *HCV data Data Set* juga terdiri dari dua tahapan yaitu *training* dan *testing*. Dimana *training* dilakukan untuk membuat sebuah model klasifikasi sedangkan *testing* dilakukan untuk menentukan *output* dari hasil klasifikasi.

4.1.2.3 Evaluating

Hasil dari proses *learning HCV data Data Set* juga akan dilakukan proses evaluasi dengan menggunakan metode *confusion matrix*. *Output* evaluasi *confusion matrix* pada *HCV data Data Set* memiliki 3 parameter keluaran yaitu *accuracy*, *sensitivity*, dan *specificity*. Dibawah ini akan ditampilkan hasil evalasi dari beberapa sekenario yang telah dilakukan pada *HCV data Data Set*.

Evaluasi akan dilakukan pada sekenario klasifikasi dengan jumlah hidden nodes yang berbeda beda di tiap trainingnya guna mendapatkan hasil klasifikasi terbaik. Hasil dari sekenario ini dapat dilihat pada Tabel 4.4.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 4 Hasil klasifikasi ELM *HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	90%	90%	88%	88%	82%	83%	85%	86%	86%
	Sensitivity	36%	45%	34%	36%	33%	31%	33%	35%	36%
	Specificity	47%	47%	37%	38%	31%	31%	39%	37%	41%

Hasil evaluasi dari beberapa ujicoba klasifikasi ELM pada Tabel 4.4, didapatkan hasil terbaik dengan menggunakan jumlah *hidden node* sebanyak 100 dengan nilai *accuracy* 90%. sedangkan untuk nilai *sensitivity* 45% dan 47% untuk nilai *specificity*.

4.2. Hasil Klasifikasi ELM Dengan SMOTE *Oversampling*

4.2.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

Pada tahapan ini juga dilakukan proses klasifikasi ELM sama seperti sebelumnya, bedanya sebelum proses klasifikasi dilakukan, *Hepatitis C Virus (HCV) for Egyptian patients Data Set* terlebih dahulu dilakukan *oversampling* dengan metode SMOTE.

SMOTE adalah metode yang digunakan untuk menyeimbangkan jumlah *class data* dengan harapan ketika jumlah *class* seimbang akan memiliki *accuracy* yang lebih baik dari sebelumnya. berikut adalah *peseudocode* dari algoritma SMOTE.

```
algoritma SMOTE

1. pilih minority class input vector x1 dan x2
2. dimana setiap sample memiliki n atribut.
3. hitung selisih pada atribut ke-i dengan rumus x12i = rand[0,1] *
dif_i
4. nilai final dari synthetic sample x1dan x2 dapat dihitung
dengan rumus x12 = rand[0,1] * dif
5. ulangi langkah tersebut sampai dataset menjadi balance atau
seimbang
```

Jumlah *dataset* yang digunakan berjumlah 1385 yang terdiri dari 362 data yang memiliki label 4, 355 data berlabel 3, 336 data berlabel 2, 332 dan berlabel 1. Seperti Gambar 4.1.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Gambar 4. 1 Jumlah Atribut *Hepatitis C Virus (HCV) for Egyptian patients Data Set* Sebelum Implementasi SMOTE

Setelah menerapan teknik SMOTE, keseluruhan *dataset* untuk klasifikasi menjadi 1448 data dengan jumlah tiap label menjadi seimbang dimana masing masing label memiliki jumlah data sebanyak 362. Proporsi jumlah data ditunjukkan pada Gambar 4.2.

Gambar 4. 2 Jumlah Atribut *Hepatitis C Virus (HCV) for Egyptian patients Data Set* Setelah Implementasi SMOTE

Gambar 4.2 meunjukan jumlah *class* menjadi seimbang dengan jumlah yang sama yaitu 362.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Proses klasifikasi *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang sudah dilakukan *oversampling* dengan metode SMOTE *Oversampling* juga dilakukan uji coba dengan menggunakan beberapa *hidden node* yang berbeda beda. Hasil beberapa uji coba tersebut dapat dilihat pada Tabel 4.5.

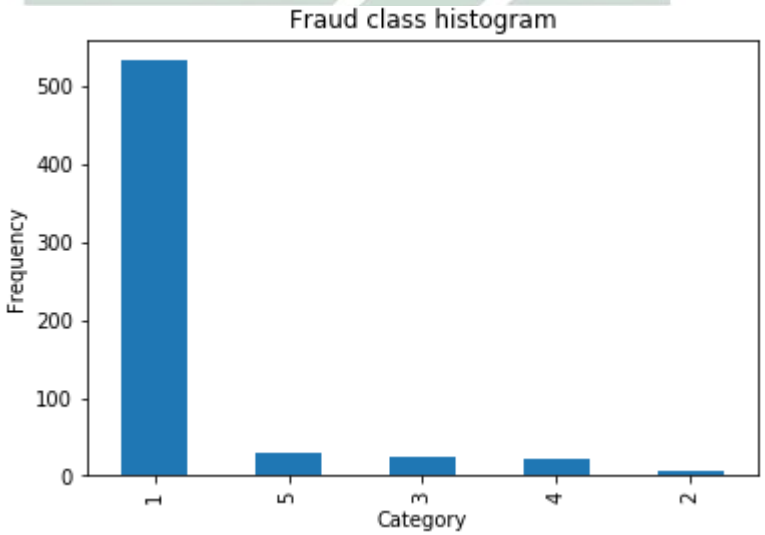
Tabel 4. 5 Hasil klasifikasi ELM dengan SMOTE *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	27%	27%	29%	27%	28%	27%	27%	27%	30%
	Sensitivity	27%	27%	29%	27%	28%	27%	27%	27%	30%
	Specificity	27%	27%	28%	27%	28%	27%	27%	27%	30%

Dari hasil pada Tabel 4.5 diketahui bahwa *accuracy* terabaik didapatkan pada sekenario pelatihan yang menggunakan *hidden node* sebanyak 150 Dengan nilai *accuracy* 30% selain *accuracy* , nilai *sensitivity* dan *specificity* dari percobaan dengan menggunakan 1500 hidden node juga memiliki nilai yang sama yaitu 30%.

4.2.2 *HCV data Data*

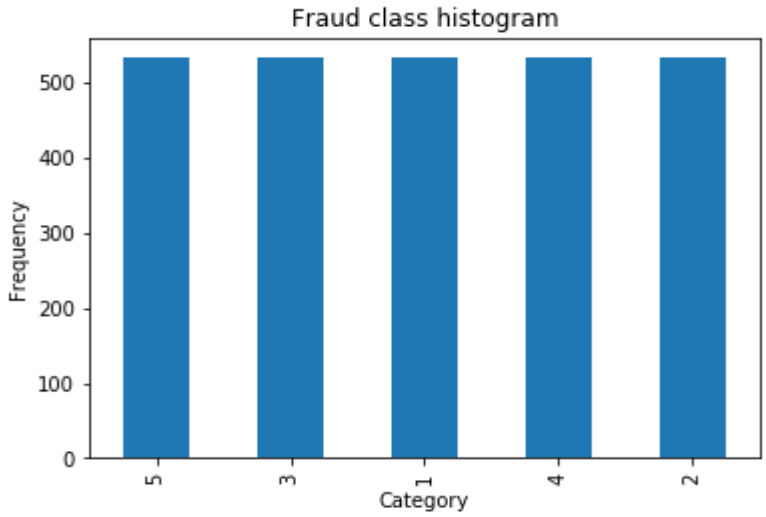
Sama seperti *Hepatitis C Virus (HCV) for Egyptian patients Data Set*, *Oversampling* juga dilakukan pada *HCV data Data Set*. Jumlah *dataset* yang digunakan berjumlah 615 yang terdiri dari 533 data yang memiliki label 1, 30 data berlabel 2, 24 data berlabel 3, 7 data berlabel 5, 21 dan berlabel 4. Seperti Gambar 4.3.



Gambar 4. 3 Jumlah Atribut *HCV data Data Set* Sebelum Implementasi SMOTE

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Setelah menerapan teknik SMOTE, keseluruhan *dataset* untuk klasifikasi menjadi 2665 data dengan jumlah tiap label menjadi seimbang dimana masing masing label memiliki jumlah data sebanyak 533. Proporsi jumlah data ditunjukkan pada Gambar 4.4.



Gambar 4. 4 Jumlah Atribut *HCV data Data Set* Setelah implementasi SMOTE

Gambar 4.4 meunjukan jumlah class *dataset* menjadi seimbang dengan jumlah yang sama yaitu 533.

Proses klasifikasi *HCV data Data Set* yang sudah dilakukan *oversampling* dengan metode SMOTE *Oversampling* juga dilakukan beberapa uji coba dengan menggunakan *hidden node* yang berbeda dengan evaluasi menggunakan *confusion matrix* di setiap uji coba. Hasil evaluasi dari uji coba tersebut dapat dilihat pada Tabel 4.6.

Tabel 4. 6 Hasil Klasifikasi ELM Dengan SMOTE *HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	98%	98%	98%	98%	98%	97%	91%
	Sensitivity	92%	96%	98%	98%	98%	98%	98%	97%	91%
	Specificity	93%	96%	98%	98%	98%	98%	98%	97%	92%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Pada Tabel 4.6, hasil terbaik didapatkan dengan pengunaan beberapa *hidden node* yaitu 250, 300, 500, 750, dan 900 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 90%.

4.3. Hasil Klasifikasi ELM Dengan *Principal Component Analysis*

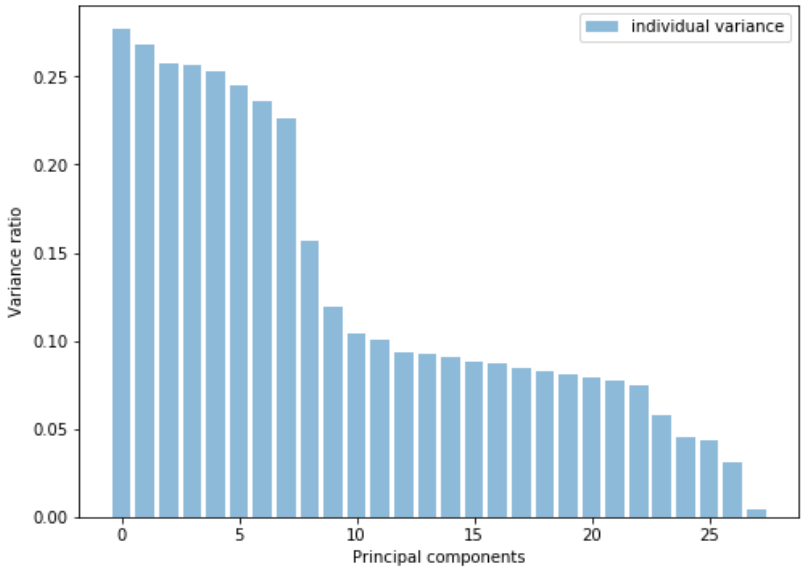
4.3.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

PCA menemukan set dimensi baru (atau set *basis of view*) sedemikian rupa sehingga semua dimensi bersifat ortogonal dan diberi peringkat sesuai dengan varian data di dalamnya. Algoritma PCA dijabarkan dalam bentuk *pseudocode* dibawah ini.

```
algoritma PCA

1. buat matrix dari dataset X
2. Hitung covariance matrix X dari dataset
3. Hitung nilai eigen vectors dan corresponding eigen values
4. Urutkan nilai eigen vectors menurut nilai eigen values dari
yang terbesar hingga terkecil
5. pilih k eigen vector pertama yang akan menjadi k dimension
baru
6. Transform  original n dimensional data points menjadi k
dimensions
```

Dengan metode PCA komponen atau atribut mewakili dimensi akan terlihat mana yang memproyeksikan nilai yang lebih rendah dan yang lebih tinggi.



Gambar 4. 5 Diagram Hasil PCA *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dari keluaran Gambar 4.5 memproyeksikan tinggi rendahnya nilai informasi yang dimiliki oleh suatu atribut. Dari hasil tersebut dapat ditentukan jumlah *n-component* yang akan digunakan untuk proses ELM

Proses klasifikasi *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang sebelumnya dilakukan seleksi fitur dengan menggunakan algoritma *principal component analysis*, akan dilakukan beberapa percobaan dengan menggunakan jumlah *n-component* yang berbeda beda dengan menggunakan hidden node yang berbeda juga pada setiap percobaan *n-component*. Hal ini dilaukan untuk mendapatkan hasil evaluasi terbaik dari proses klasifikasi.

a. Menggunakan *n-component* Berjumlah 25

Menggunakan 25 *n-component* menunjukan bahwa algoritma *principal component analysis* telah mereduksi 28 fitur *dataset* Mesir menjadi 25 *Principal Component* . Hasil dari klasfikasi dengan 25 *n – component* dapat dilihat pada Tabel 4.7.

Tabel 4. 7 Hasil Klasifikasi PCA+ELM Dengan 25 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	25%	26%	25%	25%	25%	25%	26%	28%	27%
	Sensitivity	25%	26%	25%	25%	25%	25%	26%	28%	27%
	Specificity	25%	26%	25%	25%	25%	25%	27%	28%	27%

Dari berapa uji coba yang dilakukan, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 100 Dengan nilai *accuracy*. *sensitivity*, dan *specificity* yang sama yaitu 28%.

b. Menggunakan *n-component* Berjumlah 20

Selanjutnya jumlah *n-component* yang akan digunakan adalah 20 yang menunjukan bahwa algoritma *principal component analysis* telah mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menjadi 20 *Principal Component* . Hasil uji coba dapat dilihat pada Tabel 4.8.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 8 Hasil Klasifikasi PCA+ELM Dengan 20 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	25%	24%	26%	25%	25%	24%	26%	27%
	Sensitivity	26%	25%	24%	26%	25%	25%	24%	26%	27%
	Specificity	26%	25%	24%	26%	25%	25%	24%	26%	27%

Dari berapa uji coba yang dilakukan, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 150 Dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 27%.

c. Menggunakan *n-component* Berjumlah 17

Sekenario selanjutnya, algoritma *principal component analysis* akan mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menjadi 17 *Principal Component* yang akan digunakan pada proses klasifikasi. Hasil dari uji coba dapat dilihat pada Tabel 4.9.

Tabel 4. 9 Hasil Klasifikasi PCA+ELM Dengan 17 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	26%	25%	25%	25%	24%	26%	27%	26%
	Sensitivity	26%	26%	25%	25%	25%	24%	26%	27%	26%
	Specificity	26%	25%	25%	25%	25%	24%	26%	27%	26%

Dari beberapa uji coba, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 1000 dengan nilai *accuracy* sebesar 27%. Selain *accuracy*, nilai *sensitivity* dan *specificity* juga memiliki hasil yang sama yaitu 27%.

d. Menggunakan *n-component* Berjumlah 15

Pada sekenario ini algoritma *principal component analysis* akan mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menjadi 15 *Principal Component* dan akan dilanjutkan untuk proses klasifikasi dimana hasil dari uji coba yang dilakukan dapat dilihat pada Tabel 4.10.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 10 Hasil Klasifikasi PCA+ELM Dengan 15 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	25%	26%	25%	26%	26%	26%	24%	28%	26%
	Sensitivity	25%	26%	25%	25%	26%	26%	24%	29%	25%
	Specificity	24%	25%	24%	26%	26%	26%	24%	28%	25%

Dari Tabel 4.10 dapat diketahui bahwa hasil terbaik didapatkan dengan penggunaan *hidden node* berjumlah 1000 dengan nilai *accuracy* dan *specificity* sebesar 28%. sedangkan untuk nilai *sensitivity* lebih besar yaitu 29%.

e. Menggunakan *n-component* Berjumlah 12

Selanjutnya, jumlah *n-componet* yang akan digunakan adalah 12. Dimana algoritma *principal component analysis* telah mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menjadi 12 *Principal Component* yang akan digunakan untuk proses klasifikasi. Hasil dari uji coba dapat dilhat pada Tabel 4.11.

Tabel 4. 11 Hasil Klasifikasi PCA+ELM Dengan 12 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	25%	26%	25%	26%	26%	24%	25%	25%
	Sensitivity	26%	25%	26%	25%	26%	26%	24%	25%	25%
	Specificity	26%	25%	26%	25%	26%	26%	24%	24%	25%

Dari Tabel 4.11 dapat diketahui bahwa hasil terbaik didapatkan dengan penggunaan *hidden node* berjumlah 1000 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 26%.

f. Menggunakan *n-component* Berjumlah 10

Terakhir , menggunakan 10 *n-component* yang berarti algoritma *principal component analysis* akan mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menjadi 10 *Principal Component* yang selanjutnya akan dilakukan klasifikasi. hasil dari uji coba dapat dilihat pada Tabel 4.12.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

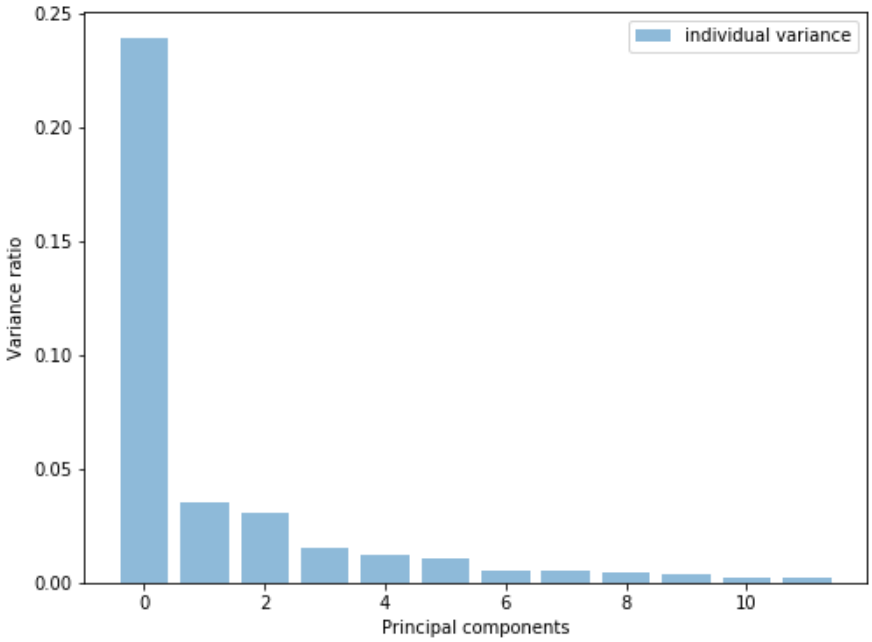
Tabel 4. 12 Hasil Klasifikasi PCA+ELM dengan 10 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	25%	24%	24%	24%	25%	24%	25%	25%	25%
	Sensitivity	25%	24%	24%	24%	25%	24%	25%	25%	25%
	Specificity	2%5	24%	24%	23%	25%	24%	25%	25%	25%

Dari Tabel 4.12 dapat diketahui bahwa hasil terbaik didapatkan dengan penggunaan beberapa *hidden node* dengan jumlah 50, 500, 900,1000, dan 1500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 25%.

4.3.2 *HCV data Data Set*

Tahapan featured selection HCV data Data Set juga dilakukan dengan menggunakan algoritma yang sama dengan *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yaitu *principal component analys*. Dengan metode PCA komponen atau atribut mewakili dimensi akan terlihat mana yang memproyeksikan nilai yang lebih rendah dan yang lebih tinggi pada *HCV data Data Set*.



Gambar 4. 6 Diagram Hasil PCA *HCV data Data Set*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dari Gambar 4.6 memproyeksikan tinggi rendahnya nilai informasi yang dimiliki oleh suatu atribut. Dari hasil tersebut dapat ditentukan jumlah *n-component* yang akan digunakan untuk proses klasifikasi.

HCV data Data Set yang sebelumnya sudah dilakukan tahapan seleksi fitur dengan menggunakan algoritma *principal component analysis* selanjutnya akan dilakukan klasifikasi dengan algoritma *extreme leraning machine*, pada proses kalsifikasi tersebut akan dilakukan beberapa skenario percobaan dengan menggunakan jumlah *n-component* yang berbeda beda dengan menggunakan *hidden node* yang berbeda juga pada setiap percobaan n-component. Pemilihan skenario ini dilakukan untuk mendapatkan hasil maksimal atau terbaik dari implementasi metode *principal component analysis.*

a. Menggunakan *n-component* Berjumlah 10

Pada skenario ini jumlah *n-componet* yang akan digunakan adalah 10. Dimana algoritma *principal component analysis* telah mereduksi 12 fitur *HCV data Data Set* menjadi 10 *Principal Component* yang akan digunakan untuk proses klasifikasi. Hasil dari uji coba dapat dilhat pada Tabel 4.13.

Tabel 4. 13 Hasil Klasifikasi PCA+ELM Dengan 10 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	91%	91%	90%	88%	84%	84%	86%	86%	87%
	Sensitivity	48%	48%	43%	35%	35%	30%	36%	36%	36%
	Specificity	71%	57%	62%	38%	37%	34%	41%	42%	41%

Dari hasil uji coba tersebut hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 50 dengan nilai *accuracy* sebesar 91%. Sedangkan untuk nilai *sensitivity* 48% dan nilai *pecificity* sebesar 71%.

b. Menggunakan *n-component* Berjumlah 8

Selanjutnya jumlah *n-component* yang akan digunakan adalah 8 yang menunjukan bahwa algoritma *principal component analysis* telah mereduksi 12 fitur *HCV data Data Set* menjadi 8 *Principal Component* . Hasil uji coba dapat dilihat pada Tabel 4.14.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 14 Hasil Klasifikasi PCA+ELM dengan 8 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	91%	89%	88%	83%	85%	86%	86%	86%
	Sensitivity	48%	48%	37%	37%	31%	32%	34%	34%	31%
	Specificity	56%	64%	39%	42%	39%	33%	42%	40%	49%

Dari hasil uji coba tersebut, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 50 dengan nilai *accuracy* sebesar 92%. Sedangkan untuk nilai *sensitivity* 48% dan nilai *specificity* sebesar 56%.

c. Menggunakan *n-component* Berjumlah 6

Pada sekenario ini algoritma *principal component analysis* akan mereduksi 12 fitur *HCV data Data Set* menjadi 6 *Principal Component* dan akan dilanjutkan untuk proses klasifikasi dimana hasil dari uji coba yang dilakukan dapat dilihat pada Tabel 4.15.

Tabel 4. 15 Hasil Klasifikasi PCA+ELM Dengan 6 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	92%	89%	88%	83%	85%	85%	85%	86%
	Sensitivity	48%	49%	40%	36%	31%	33%	34%	33%	40%
	Specificity	62%	66%	45%	48%	39%	37%	35%	37%	65%

Dari hasil uji coba Tabel 4.15 hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 100 dengan nilai *accuracy* sebesar 92%. Sedangkan untuk nilai *sensitivity* 49% dan nilai *specificity* sebesar 66%.

d. Menggunakan *n-component* Berjumlah 5

Pada sekenario ini algoritma *principal component analysis* akan mereduksi 12 fitur *HCV data Data Set* menjadi 5 *Principal Component* dan akan dilanjutkan untuk proses klasifikasi dimana hasil dari uji coba yang dilakukan dapat dilihat pada Tabel 4.16.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 16 Hasil Klasifikasi PCA+ELM Dengan 5 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	91%	91%	88%	87%	84%	84%	85%	85%	85%
	Sensitivity	43%	45%	35%	36%	37%	34%	34%	34%	35%
	Specificity	67%	48%	39%	39%	39%	35%	38%	36%	40%

Dari hasil uji coba tersebut, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 50 dengan nilai *accuracy* sebesar 91%. Sedangkan untuk nilai *sensitivity* 43% dan nilai *specificity* sebesar 67%.

e. Menggunakan *n-component* Berjumlah 3

Terakhir , menggunakan 3 *n-component* yang berarti algoritma *principal component analysis* akan mereduksi 12 fitur *HCV data Data Set* menjadi 3 *Principal Component* yang selanjutnya akan dilakukan klasifikasi. hasil dari uji coba dapat dilihat pada Tabel 4.17.

Tabel 4. 17 Hasil Klasifikasi PCA+ELM Dengan 3 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	91%	92%	88%	87%	84%	84%	85%	84%	86%
	Sensitivity	44%	46%	36%	36%	35%	33%	35%	29%	35%
	Specificity	72%	67%	39%	38%	38%	32%	39%	29%	44%

Dari hasil uji coba tersebut, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 100 dengan nilai *accuracy* sebesar 92%. Sedangkan untuk nilai *sensitivity* 46% dan nilai *specificity* sebesar 67%.

4.4. Hasil Klasifikasi ELM Dengan Kombinasi SMOTE dan PCA

4.4.1 *Hepatitis C Virus (HCV) for Egyptian patients Data Set*

Pada skenario terakhir atau skenario inti akan dilakukan kalsifikasi dengan menggunakan *dataset* yang sebelumnya dilakukan *oversampling* dengan metode SMOTE dan seleksi fitur dengan menggunakan metode *principal component analysis*. Akan dilakukan beberapa percobaan dengan menggunakan jumlah *n-component* yang berbeda beda dengan menggunakan *hidden node* yang berbeda juga pada setiap percobaan *n-component*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

a. Menggunakan *n-component* Berjumlah 25

Pada sekenario pertama, *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang sudah dilakukan *oversampling* dengan menggunakan metode SMOTE selanjutnya akan direduksi menggunakan algoritma *principal component analysis.* Dengan PCA 28 *fitur Hepatitis C Virus (HCV) for Egyptian patients Data Set* akan direduksi menjadi 25 *Principal Component* dan selanjutnya akan dilakukan klasifikasi. Hasil uji coba dapat dilihat pada Tabel 4.18.

Tabel 4. 18 Hasil Kasifikasi SMOTE+PCA+ELM Dengan 25 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	29%	29%	27%	28%	28%	28%	28%	30%
	Sensitivity	26%	29%	29%	27%	28%	28%	28%	28%	30%
	Specificity	26%	29%	29%	27%	28%	28%	28%	28%	30%

Dari Tabel 4.18 dapat diketahui bahwa hasil terbaik didapatkan dengan menggunakan *hidden node* dengan jumlah 1500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 30%.

b. Menggunakan *n-component* Berjumlah 20

Selanjutnya akan digunakan 20 *n-component* yang artinya algoritma *principal component analysis* telah mereduksi fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang telah di *oversampling* menjadi 20 *Principal Component*. Hasil uji coba dapat dilihat pada tabel 4.19.

Tabel 4. 19 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 20 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	27%	27%	29%	29%	29%	30%	29%	30%
	Sensitivity	26%	27%	27%	29%	29%	29%	30%	29%	30%
	Specificity	26%	27%	26%	29%	29%	29%	30%	29%	30%

Hasil dari ujicoba tersebut menunjukkan, hasil terbaik didapatkan dengan menggunakan *hidden node* sebanyak 1500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang sama yaitu 30%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

c. Menggunakan *n-component* Berjumlah 17

Pada sekenario ini jumlah *n-component* yang digunakan adalah 17 yang menunjukan bahwa algoritma *principal component analysis* telah mereduksi fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang telah di *oversampling* menjadi 17 *Principal Component*. Untuk hasil ujicoba bisa dilihat pada Tabel 4.20.

Tabel 4. 20 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 17 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	27%	27%	29%	29%	29%	30%	29%	30%
	Sensitivity	26%	27%	27%	29%	29%	29%	30%	29%	30%
	Specificity	26%	27%	26%	29%	29%	29%	30%	29%	30%

Tabel 4.20 menunjukkan bahwa hasil terbaik didapatkan dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 30% hasil tersebut didapatkan dengan menggunakan 1500 *hidden node*.

d. Menggunakan *n-component* Berjumlah 15

Sekenario selanjutnya akan mengunakan *n-component* berjumlah 15. Dimana dari 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* direduksi menjadi 15 *Principal Component* oleh algoritma *principal component analysis.* Untuk hasil uji coba dapat dilihat pada Tabel 4.21.

Tabel 4. 21 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 15 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	26%	28%	26%	27%	28%	30%	28%	30%	30%
	Sensitivity	26%	28%	26%	27%	28%	30%	28%	30%	30%
	Specificity	26%	28%	26%	27%	28%	30%	28%	30%	30%

Dari hasil uji coba tersebut, hasil terbaik didapatkan dengan jumlah *hidden node* sebanyak 1000 dan 1500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 30%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

e. Menggunakan *n-component* Berjumlah 12

Sekenario ini akan menggunakan 12 *n-component* dari hasil reduksi algoritma *principal component analysis*. Dimana dari 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set* direduksi menjadi 12 *Principal Component*. Untuk hasil dari sekenario ini dapat dilihat pada Tabel 4.22.

Tabel 4. 22 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 12 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	27%	27%	28%	28%	30%	28%	31%	30%	30%
	Sensitivity	27%	27%	28%	28%	30%	28%	31%	30%	30%
	Specificity	27%	27%	28%	28%	30%	28%	31%	30%	30%

Tabel 4.22 menunjukan bahwa dengan jumlah *hidden node* sebanyak 900 didapatkan hasil terbaik dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 31%.

f. Menggunakan *n-component* Berjumlah 10

Sekenario terakhir akan menggunakna 10 *n-component* yang didapatkan dari mereduksi 28 fitur *Hepatitis C Virus (HCV) for Egyptian patients Data Set*. 10 fitur ini akan dilakukan klasifikasi dengan menggunakan beberapa *hidden node* yang berbeda. Hasil dari skeneario ini dapat dilihat pada Tabel 4.23.

Tabel 4. 23 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 10 *n-component Hepatitis C Virus (HCV) for Egyptian patients Data Set dataset Mesir*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	27%	26%	28%	28%	29%	28%	30%	29%	31%
	Sensitivity	27%	26%	28%	28%	29%	28%	30%	29%	31%
	Specificity	27%	26%	28%	28%	29%	28%	30%	29%	31%

Dari Tabel 4.23 menunjukan bahwa hasil terbaik dari serangkaian skenario didapatkan dengan jumlah *hidden node* sebanyak 1500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 31%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.4.2 *HCV data Data Set*

Pada skenario terakhir proses klasifikasi *HCV data Data Set* juga akan dilakukan kalsifikasi dengan menggunakan *dataset* yang sebelumnya dilakukan *oversampling* dengan metode SMOTE dan seleksi fitur dengan menggunakan metode *principal component analysis*. Beberapa skenario percobaan akan dilakukan dengan menggunakan jumlah *n-component* dan *hidden node* yang berbeda untuk mendapatkan hasil evaluasi terbaik.

a. Menggunakan *n-component* Berjumlah 10

Pada skenario pertama, *HCV data Data Set* yang sudah dilakukan *oversampling* dengan menggunakan metode SMOTE selanjutnya akan direduksi menggunakan algoritma *principal component analysis.* Dengan metode PCA, 12 fitur *HCV data Data Set* akan direduksi menjadi 10 *Principal Component* dan selanjutnya akan dilakukan klasifikasi. Hasil uji coba dapat dilihat pada Tabel 4.24.

Tabel 4. 24 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 10 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	98%	99%	98%	98%	97%	97%	91%
	Sensitivity	92%	96%	98%	99%	98%	98%	97%	97%	91%
	Specificity	92%	96%	98%	99%	98%	98%	97%	97%	92%

Dari hasil uji coba didapatkan, hasil terbaik didapatkan dengan jumlah *hidden node* 250, 300, 500, dan 750 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 99%.

b. Menggunakan *n-component* Berjumlah 8

Selanjutnya akan digunakan 8 *n-component* yang artinya algoritma *principal component analysis* telah mereduksi fitur *HCV data Data Set* yang telah dilakukan *oversampling* menjadi 8 *Principal Component*. Hasil uji coba dapat dilihat pada Tabel 4.25.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 25 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 8 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	98%	99%	99%	98%	98%	96%	90%
	Sensitivity	92%	96%	98%	99%	99%	98%	98%	97%	90%
	Specificity	92%	96%	98%	99%	99%	98%	98%	97%	91%

Dari hasil uji coba, didapatkan hasil terbaik dengan jumlah *hidden node* 250, 300, dan 500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 99%.

c. Menggunakan *n-component* Berjumlah 6

Sekenario ini akan menggunakan 12 *n-component* dari hasil reduksi algoritma *principal component analysis*. Dimana dari 12 fitur *HCV data Data Set* direduksi menjadi 6 *Principal Component*. Untuk hasil dari skenario ini dapat dilihat pada Tabel 4.26.

Tabel 4. 26 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 6 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	98%	99%	99%	98%	98%	97%	91%
	Sensitivity	92%	96%	98%	99%	99%	98%	98%	97%	91%
	Specificity	92%	96%	98%	99%	99%	98%	98%	97%	91%

Tabel 4,26 menunjukan bahwa dengan jumlah *hidden node* sebanyak 900 dan 300 didapatkan hasil terbaik dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 99%.

d. Menggunakan *n-component* Berjumlah 5

Selanjutnya jumlah *n-component* yang akan digunakan adalah 5 yang menunjukan bahwa algoritma *principal component analysis* telah mereduksi 12 fitur *HCV data Data Set* menjadi 5 *Principal Component* . Hasil uji coba dapat dilihat pada Tabel 4.27.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 27 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 5 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	99%	99%	99%	99%	98%	97%	90%
	Sensitivity	92%	96%	99%	99%	99%	99%	98%	97%	90%
	Specificity	93%	96%	99%	99%	99%	99%	98%	97%	92%

Dari hasil uji coba tersebut, menunjukkan bahwa hasil terbaik dari serangkaian skenario didapatkan dengan jumlah *hidden node* 250, 300, 500, dan 750 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 99%.

e. Menggunakan *n-component* Berjumlah 3

Sekenario terakhir akan menggunakna 3 *n-component* yang didapatkan dari mereduksi 12 fitur *HCV data Data Set.* 3 fitur ini akan dilakukan klasifikasi dengan menggunakan beberapa *hidden node* yang berbeda. Hasil dari skeneario ini dapat dilihat pada Tabel 4.28.

Tabel 4. 28 Hasil Klasifikasi SMOTE+PCA+ELM Dengan 8 *n-component HCV data Data Set*

		Hidden Nodes								
		50	**100**	**250**	**300**	**500**	**750**	**900**	**1000**	**1500**
Validasi	**Accuracy**	92%	96%	99%	99%	99%	98%	98%	97%	90%
	Sensitivity	92%	96%	99%	99%	99%	98%	98%	97%	90%
	Specificity	92%	96%	99%	99%	99%	98%	98%	97%	91%

Dari Tabel 4.28 menunjukan bahwa hasil terbaik dari serangkaian skenario didapatkan dengan jumlah *hidden node* 250, 300, dan 500 dengan nilai *accuracy*, *sensitivity*, dan *specificity* yang berjumlah sama yaitu 99%.

4.5. Analisis Hasil

Bagian ini akan membahas tentang hasil penelitian yang telah dipaparkan untuk menjawab rumusan permasalahan yang telah maupun sedang diteliti. Sehingga akan ditemukan hal baru yang bisa dikembangkan lagi pada penelitian kedepannya untuk mendapatkan hasil yang lebih baik lagi.

Skenario pertama pada penelitian ini adalah implementasi ELM pada dua *dataset* yang berbeda. Aktivasi fungsi yang digunakan adalah sigmoid. Pada

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sekenario ini beberapa uji coba dilakukan dengan menggunakan *hidden node* yang berbeda beda. Hasil terbaik dari serangkaian uji coba kedua *dataset* akan dibandingkan seperti pada Tabel 4.29.

Tabel 4. 29 Perbandingan Hasil Klasifikasi Metode ELM

		Hepatitis C Virus (HCV) for Egyptian patients Data Set	*HCV data Data Set*
Validasi	Accuracy	26%	90%
	Sensitivity	26%	45%
	Specificity	26%	47%

Dari perbandingan diatas menunjukan bahwa *HCV data Data Set* memiliki hasil klasifikasi yang lebih baik dengan nilai *accuracy* sebesar 90% dibandingkan *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang hanya mendapatkan nilai *accuracy* 26%. meskipun mendapatkan *accuracy* yang tinggi , nilai *sensitivity* dan *specificity HCV data Data Set* sangat kecil yaitu 45% dan 47% sedangkan *Hepatitis C Virus (HCV) for Egyptian patients Data Set* memiliki nilai *sensitivity* dan *specificity* yang sama yaitu 26%.

Selanjutnya pada skenario berikutnya klasifikasi metode ELM akan dilakukan dengan menggunakan metode *oversampling* SMOTE yang bertujuan untuk menyeimbangkan jumlah *class* data dengan membuat sample baru. Hasil terbaik dari serangkaian uji coba kedua *dataset* terlihat seperti pada Tabel 4.30.

Tabel 4. 30 Perbandingan Hasil Klasifikasi ELM dengan SMOTE

		Hepatitis C Virus (HCV) for Egyptian patients Data Set tanpa SMOTE	*Hepatitis C Virus (HCV) for Egyptian patients Data Set* dengan SMOTE	*HCV data Data Set* tanpa SMOTE	*HCV data Data Set* dengan SMOTE
Validasi	Accuracy	26%	30%	90%	99%
	Sensitivity	26%	30%	45%	99%
	Specificity	26%	30%	47%	99%

Pada *HCV data Data Set* metode SMOTE sangat memberikan pengaruh peningkatan hasil evaluasi. Pada sisi *accuracy* terdapat peningkatan dari 90%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menjadi 98%. Selain *accuracy* hal yang paling menonjol adalah peningkatan nilai *sensitivity* dan *specificity*, dimana sebelum *oversampling* nilai yang dihasilkan sangat kecil yaitu 45% dan 47%. Setelah di *oversampling* nilai *sensitivity* dan *specificity* sama sama menjadi 98%, hal ini menunjukan bahwa data *imbalanced class* memiliki nilai *sensitivity* dan *specificity* yang kecil meskipun nilai *accuracy* cukup tinggi dan permasalahan seperti ini bisa diatasi dengan metode *oversampling* seperti SMOTE.

Sedangkan pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* SMOTE juga memberikan pengaruh dimana nilai *accuracy* , *sensitivity* dan *specificity* yang awalnya 26% meningkat menjadi 30%.

Pada skenario ketiga akan dilakukan klasifikasi motde ELM dengan menggunakan metode *feature selection* PCA. Hasil terbaik dari serangkaian uji coba kedua *dataset* yang telah diimplementasikan PCA dapat dilihat pada Tabel 4.31.

Tabel 4. 31 Perbandingan Hasil Klasifikasi ELM dengan PCA

		Hepatitis C Virus (HCV) for Egyptian patients Data Set tanpa PCA	*Hepatitis C Virus (HCV) for Egyptian patients Data Set* dengan PCA	*HCV data Data Set* tanpa PCA	*HCV data Data Set* dengan PCA
Validasi	Accuracy	26%	28%	90%	92%
	Sensitivity	26%	29%	45%	49%
	Specificity	26%	28%	47%	66%

Dari Tabel 4.31 diatas dapat dilihat bahwa metode PCA mampu meningkatkan hasil validasi dari kedua *dataset*. Pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* nilai *accuracy* dan nilai *specificity* meningkat dari 26% menjadi 28% dan untuk nilai *sensitivity* meningkat dari 26% menjadi 29%. *HCV data Data Set* juga mengalami peningkatan hasil evaluasi dimana nilai *accuracy* meningkat dari 90% menjadi 92% dan untuk nilai *sensitivity* dan *specificity* meningkat dari 45% dan 47% menjadi 49% dan 66%.

Dan pada skenario terakhir kedua metode tersebut dikombinasikan pada kedua *dataset* yang selanjutnya akan dilakukan klasifikasi menggunakan metode

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ELM. Hasil terbaik dari serangkaian uji coba kedua *dataset* terlihat seperti pada Tabel 4.32.

Tabel 4. 32 Perbandingan Hasil Klasifikasi ELM Dengan Kombinasi SMOTE + PCA

		Hepatitis C Virus (HCV) for Egyptian patients Data Set tanpa SMOTE+PCA	*Hepatitis C Virus (HCV) for Egyptian patients Data Set* dengan SMOTE+PCA	*HCV data Data Set* tanpa SMOTE+PCA	*HCV data Data Set* dengan SMOTE+PCA
Validasi	Accuracy	26%	31%	90%	99%
	Sensitivity	26%	31%	45%	99%
	Specificity	26%	31%	47%	99%

Dari hasil Tabel 4.32 diatas dapat disimpulkan bahwa kombinasi SMOTE dam PCA juga mampu meningkatkan hasil evaluasi dari kedua *dataset*. Hasil kombinasi ini dapat dikatakan lebih baik dari pada hanya menerapkan metode SMOTE atau PCA saja, perbandingan tersebut dapat dilihat pada Tabel 4.33.

Tabel 4. 33 Perbandingan Hasil klasifikasi Kedua Dataset

		ELM	**ELM + SMOTE**	**ELM + PCA**	**ELM+ SMOTE + PCA**
HCV data Data Set	Accuracy	26%	30%	28%	31%
	Sensitivity	26%	30%	28%	31%
	Specificity	26%	30%	28%	31%
Hepatitis C Virus (HCV) for Egyptian patients Data Set	Accuracy	90%	98%	92%	99%
	Sensitivity	45%	98%	49%	99%
	Specificity	47%	98%	66%	99%

Dari tabel 4.33 menunjukkan bahwa kombinasi metode SMOTE dan PCA memberikan hasil yang jauh lebih baik. Pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* evaluasi klasifikasi dengan *metode extreme leraning machine* menghasilan nilai *accuracy sensitivity*, dan *specificity* sebesar 26%. Dengan kombinasi metode SMOTE dan PCA *nilai accuracy sensitivity*, dan *specificity*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

meningkat menjadi 31%, lebih besar daripada jika hanya penerapan SMOTE saja yang nilai *accuracy sensitivity*, dan *specificity* sebesar 30% Begitu juga dengan penerapan metode PCA saja yang nilai *accuracy sensitivity*, dan *specificity* sebesar 28%.

Dan pada *HCV data Data Set* kombinasi metode SMOTE dan PCA juga mampu meningkatkan hasil klasifikasi dengan nilai *accuracy sensitivity*, dan *specificity* yang sama yaitu 99%. Hasil tersebut lebih baik dibanding hanya penerapan metode SMOTE saja dimana nilai *accuracy sensitivity*, dan *specificity* sebesar 98%. dan penerapan metode PCA saja dengan nilai *accuracy sensitivity*, dan *specificity* berturut turut adalah 92%, 49%, dan 66%.

Akan tetapi terdapat perbedaan yang cukup besar pada hasil klasifikasi kedua *dataset* dimana nilai *accuracy sensitivity*, dan *specificity Hepatitis C Virus (HCV) for Egyptian patients Data Set* sangat kecil apabila dibandingkan dengan *HCV data Data Set*. Selain hasil evaluasi yang kecil , nilai korelasi antar atribut pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* juga sangat rendah apabila dibandingkan dengan *HCV data Data Set*, Seperti yang terlihat pada Tabel 4.34.

Tabel 4.34 Perbandingan 7 atribut dengan korelasi terbaik

Atribut *Hepatitis C Virus (HCV) for Egyptian patients Data Set*	Nilai korelasi dengan label	Nilai korelasi dengan label	Atribut *HCV data Data Set*
Jaundice	0.020	0.028	ALP
RNA Base	0.029	0.106	Age
RNA EF	0.030	0.106	ALT
ALT after 24 w	0.034	0.182	CREA
RNA 12	0.034	0.471	GGT
ALT 1	0.037	0.473	Bilirubin (BIL)
Nausea/Vomting	0.055	0.648	AST

Tabel 4.34 merupakan hasil analisis menggunakan metode EDA dimana atribut yang dibandingkan diatas merupakan 7 atribut masing masing *dataset* yang memiliki nilai korelasi terbaik. Dapat dilihat bahwa nilai korelasi *Hepatitis C Virus (HCV) for Egyptian patients Data Set* sangat kecil, bahkan nilai koreasi tertinggi hanya 0.055 sedangkan nilai korelasi atribut *HCV data Data Set* cukup stabil dengan nilai tertinggi 0.648. Dari hal tersebut dapat disimpulkan bahwa hasil

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

evaluasi proses klasifikasi dari *Hepatitis C Virus (HCV) for Egyptian patients Data Set* yang kecil disebabkan oleh nilai korelasi antar atributnya yang kecil.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V
Kesimpulan dan Saran

5.1. Kesimpulan

1. Performa klasifikasi *Hepatitis C Virus (HCV) for Egyptian patients Data Set* menggunakan algoritma ELM dari beberapa percobaan menghasilkan nilai *accuracy*, *sensitivity*, dan *specificity* terbaik masing-masing 26%. Sedangkan *HCV data Data Set* menghasilkan nilai evaluasi yang cukup tinggi dengan tingkat *accuracy* sebesar 90%, untuk *sensitivity* mendapatkan nilai sebesar 45%, dan *specificity* sebesar 47%,.
2. Penerapan algoritma SMOTE pada *oversampling* terbukti dapat menangani masalah *imbalanced class* sehingga *dataset* memiliki jumlah *class label* yang seimbang. Pada riset ini, *HCV data Data Set* memiliki *imbalanced class* yang cukup besar dibandingkan dengan *Hepatitis C Virus (HCV) for Egyptian patients Data Set*. Implementasi SMOTE memberikan dampak yang cukup signifikan pada *dataset* tersebut. Tingkat *accuracy* yang awalnya 90%, meningkat menjadi 98%, nilai *sensitivity* juga meningkat dari 45% menjadi 98%, dan nilai specificity meningkat 51%, yang awalnya 57%, menjadi 98%. Begitu pula pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* mengalami peningkatan sekalipun tidak terlalu besar, dimana nilai *accuracy*, *sensitivity*, dan *specificity* yang awalnya sama yaitu 26%, semuanya meningkat menjadi 30%,
3. Metode *featured selection* PCA mampu meningkatkan hasil evaluasi dari proses klasifikasi kedua *dataset*. Pada *HCV data Data Set* nilai *accuracy* meningkat dari 90%, menjadi 92%. nilai *sensitivity* meningkat dari 45% menjadi 49%, dan untuk nilai *specificity* meningkat dari 49%, menjadi 66%. Sedangkan pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* nilai *accuracy*, *sensitivity*, dan *specificity* yang awalnya sama yaitu 26%, semuanya meningkat menjadi 28%,.
4. Bahwa kombinasi metode *oversampling* SMOTE dan *featured selection* PCA terbukti mampu meningkatkan performa klasifikasi jika dibandingkan dengan implementasi metode SMOTE maupun PCA saja. Pada *HCV data Data Set*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang nilai *accuracy, sensitivity,* dan *specificity* berturut-turut adalah 90%,, 44%,, 47%, semuanya meningkat menjadi 99%,. sedangkan pada *Hepatitis C Virus (HCV) for Egyptian patients Data Set* nilai *accuracy, sensitivity,* dan *specificity* yang awalnya sama yaitu 26%, semuanya meningkat menjadi 31%,.

5.2. Saran

Dalam penelitian Klasifikasi data HCV yang menggunakan dua *dataset* yang berbeda akan dipaparkan beberapa rekomendasi dan saran untuk perbaikan pada penelitian berikutnya.

1. Menggunaakn Metode ELM modifikasi dengan menambahkan jumlah *hidden layer* dan juga menambahkan *kernel* sehingga hasil klasifikasi akan lebih baik dari pada sebelumnya.
2. Menggunakan metode *featured selection* yang lain untuk melakukan proses seleksi fitur.
3. Menggunakan metode *oversampling* yang lain.untuk menyelesaikan permasalahan *imbalanced dataset*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

A. Jilani, T., Yasin, H., & Mohammad Yasin, M. (2011). PCA-ANN for Classification of Hepatitis-C Patients. *International Journal of Computer Applications*, *14*(7), 1–6. https://doi.org/10.5120/1899-2530

Adiwijaya, A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *Jurnal Media Informatika Budidarma*, *2*(4), 181. https://doi.org/10.30865/mib.v2i4.1043

Agustina, I. D., Anggraeni, W., & Mukhlason, A. (2005). Penerapan Metode Extreme Learning Machine untuk Peramalan Permintaan. *Digilib, Institut Teknologi Sepuluh Nopember*, 1–6.

Bataller, R., & Brenner, D. a. (2005). Science in medicine Liver fibrosis. *The Journal of Clinical Investigation*, *115*(2), 209–218. https://doi.org/10.1172/JCI200524282.The

Brataatmadja, D. (2003). Aspek Laboratorium Pada Infeksi Virus Hepatitis C. *Journal of Medicine and Health*, *3*, 13–25.

Cacoub, P., & Comarmond, C. (2017). New insights into HCV-related rheumatologic disorders: A review. *Journal of Advanced Research*, *8*(2), 89–97. https://doi.org/10.1016/j.jare.2016.07.005

Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, *340–341*, 250–261. https://doi.org/10.1016/j.ins.2016.01.033

Dewi, E. A. (2018). *Perbandingan Metode Holt Winter's Exponential Smoothing dan Extreme Learning Machine(ELM) Pada Peramalan Penjualan Semen*. *16*.

Foeady, A. Z. (2019). *Sistem klasifikasi kanker kulit berdasarkan data citra dermoscopic dengan menggunakan metode deep extreme learning machine*.

Fridayanthie, E. W. (2015). Analisa Data Minning untuk Prediksi Penyakit Hepatitis Dengan mengguakan Metode Naive Bayes dan Support Vector Machine. *JURNAL KHATULISTIWA INFORMATIKA*, *16*(1), 24–25. https://doi.org/10.30798/makuiibf.323102

Hairany, A. (2007). *Megenal bahaya hepatitis c.pdf* (pp. 2–3). pp. 2–3.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Hidayat, R. (2012). Meminimalisasi nilai error peramalan dengan algoritma extreme learning machine. *Optimasi Sistem Industri*, 187–192.

Huang, G. Bin, Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Irawan, F., & Samopa, F. (2018). A Comparative Assessment of Random Forest and SVM Algorithms , Using Combination of Principal Component Analysis and SMOTE For Accounts A Comparative Assessment of Random Forest and SVM Algorithms , Using Combination of Principal Component Analysis and SM. *The 2nd International Seminar Of Contemporary Research OnBusiness & Management*, (July).

Irawan, G. A., & Muliantara, A. (2017). Prediksi Kesuburan (Fertility) Dengan Menggunakan Principal Component Analysis Dan Klasifikasi Naive Bayes. *Jurnal Ilmiah Ilmu Komputer Universitas Udayana*, *10*(2), 10–14.

Iriadi, N. (2012). Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi. *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, *XIV*(2), 120–129.

Ismawan, F. (2015). Hasil Ekstraksi Algoritma Principal Component Analysis (PCA) untuk Pengenalan Wajah dengan Bahasa Pemograman Java Eclipse IDE. *JURNAL SISFOTEK GLOBAL*, *5*(1), 26–30.

Izzuddin, A., Wahyudi, R., & Kartini, D. P. (2018). *Pengenalan Spesies Gulma dengan Metode Ekstraksi Ciri Principal Component Analysis (PCA) dan Metode Klasifikasi Extreme Learning Machine (ELM)*. *Vol. 8*(2), 42–48.

Kasun, L. L. C., Zhou, H., Huang, G. Bin, & Vong, C. M. (2013). Representational learning with ELMs for big data. *IEEE Intelligent Systems*, *28*(6), 31–34.

Khomsah, S. (2018). Prediksi Harapan Hidup Penderita Hepatitis Kronik Menggunakan Metode-Metode Klasifikasi. *Seminar Nasional Informatika Medis*, 38–45.

Komputer, D. I., Matematika, F., Ilmu, D. A. N., & Alam, P. (2015). *Klasifikasi Fragmen Metagenom Menggunakan Principal Component Analysis Dan K-Nearest Neighbor*. (Wu 2008), 1–7.

Learning, U. M. (2019). Hepatitis C Virus (HCV) for Egyptian patients Data Set.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Https://Archive.Ics.Uci.Edu/Ml/Machine-Learning-Databases/00503/.

Learning, U. M. (2020). HCV data Data Set. *Http://Archive.Ics.Uci.Edu/Ml/Datasets/HCV+data*.

Leija, A., Reyes, J., & Rodríguez, L. (2007). Hepatic stellate cells are a major component of liver fibrosis and a target for the treatment of chronic liver disease. *Biotecnologia Aplicada*, *24*(1), 19–25.

Maldonado, S., López, J., & Vairetti, C. (2019). An alternative SMOTE oversampling strategy for high-dimensional datasets. *Applied Soft Computing Journal*, *76*, 380–389. https://doi.org/10.1016/j.asoc.2018.12.024

Mujilahwati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, *2016*(Sentika), 2089–9815.

Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, *4*(2), 151. https://doi.org/10.24076/citec.2017v4i2.106

Nasr, M., El-Bahnasy, K., Hamdy, M., & Kamal, S. M. (2018). A novel model based on non invasive methods for prediction of liver fibrosis. *ICENCO 2017 - 13th International Computer Engineering Conference: Boundless Smart Societies*, *2018-Janua*, 276–281. https://doi.org/10.1109/ICENCO.2017.8289800

Prakoso, E. C., & Wisesty, U. N. (2016). Klasifikasi Keadaan Mata Berdasarkan sinyal EEG menggunakan Extreme Learning Machines. *Indonesian Journal on Computing (Indo-JC)*, *1*(2), 97. https://doi.org/10.21108/indojc.2016.1.2.105

Pratiwi, H., & Harianto, K. (2019). Perbandingan Algoritma ELM Dan Backpropagation Terhadap Prestasi Akademik Mahasiswa. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *3*(2), 282. https://doi.org/10.30645/j-sakti.v3i2.147

Priyadarshini, R., Dash, N., & Mishra, R. (2014). A Novel approach to Predict Diabetes Mellitus using Modified Extreme Learning Machine. *2014 International Conference on Electronics and Communication Systems, ICECS 2014*. https://doi.org/10.1109/ECS.2014.6892740

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Puspaningrum, A. (2008). *Model Matematika Pada Repliaksi Virus Hepatitis C Dalam Vesicular Membrane Structure (VMS) Dengan Sistem Persamaan Diferensial*. 1–105.

Ramezankhani, A., Pournik, O., Shahrabi, J., Azizi, F., Hadaegh, F., & Khalili, D. (2016). The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of type 2 diabetes. *Medical Decision Making*, *36*(1), 137–144. https://doi.org/10.1177/0272989X14560647

Ravikumar, S. (2016). Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine. *Artificial Cells, Nanomedicine and Biotechnology*, *44*(3), 985–989. https://doi.org/10.3109/21691401.2015.1008506

Santosa, B. (2007). Data mining terapan dengan matlab. *Yogyakarta: Graha Ilmu*, 272–281.

Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri Volume*, *13*(1), 76–84. https://doi.org/10.33480/pilar.v13i1.149

Soliman, O. S., & Elhamd, A. (2014). Classification of Hepatitis C Virus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine. *International Journal of Scientific & Engineering Research*, *5*(3), 122–129. Retrieved from http://www.ijser.org

Sreejith, S., Khanna Nehemiah, H., & Kannan, A. (2020). Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection. *Computers in Biology and Medicine*, *126*(February), 103991. https://doi.org/10.1016/j.compbiomed.2020.103991

Sreekanth, M. S., Rajesh, R., & Satheeshkumar, J. (2015). Extreme learning machine for the classification of rainfall and thunderstorm. *Journal of Applied Sciences*, Vol. 15, pp. 153–156. https://doi.org/10.3923/jas.2015.153.156

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 and Learning*, *1*(3), 183–191. https://doi.org/10.23887/jere.v1i3.12016

Syakhala, A. R., Puspitaningrum, D., & Purwandari, E. P. (2015). Perbandingan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Metode Principal Component Analysis (PCA) dengan Metode Hidden Markov Model (HMM) dalam Pengenalan Identitas. *Jurnal Informatika*, *3*(2), 68–81. Retrieved from https://ejournal.unib.ac.id/index.php/rekursif/issue/archive

Utomo, C. P., & Kardiana, A. (2014). Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques. *International Journal of Advanced Research in Artificial Intelligence*, *3*(7), 10–14. https://doi.org/10.1017/CBO9781107415324.004

Wibawa, A. P. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, *3*(1), 134–138.

Wiyono, A. R., & Imah, E. M. (2018). PENGENALAN CITRA EKSPRESI WAJAH MENGGUNAKAN ALGORITMA PRINCIPAL COMPONENT ANALYSIS (PCA) DAN EXTREME LEARNING MACHINE (ELM). *Jurnal Ilmiah Matematika*, *6*(2), 2–6.

Wu, D., Wang, Z., Chen, Y., & Zhao, H. (2016). Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset. *Neurocomputing*, *190*, 35–49. https://doi.org/10.1016/j.neucom.2015.11.095

Xie, W., Li, Y., & Ma, Y. (2016). Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing*, *173*, 930–941. https://doi.org/10.1016/j.neucom.2015.08.048

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id