

**ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK
DIAGNOSIS PENYAKIT HATI (*LIVER*)**

Skripsi

Oleh

SALSABILLA JULIA FARHANA

NPM 1917051004



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRAK

ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK DIAGNOSIS PENYAKIT HATI (*LIVER*)

Oleh

SALSABILLA JULIA FARHANA

Hati (*Liver*) adalah organ dalam tubuh manusia yang memiliki peran penting. Penyakit hati (*liver*) adalah gangguan pada hati yang menyebabkan organ tersebut berhenti berfungsi dengan baik. Tujuan dari penelitian ini adalah menerapkan berbagai metode klasifikasi seperti *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* untuk mendiagnosis penyakit hati (*liver*). Penerapan metode klasifikasi ini diharapkan dapat memberikan informasi dalam mendiagnosis penyakit hati (*liver*). Penelitian ini menggunakan *dataset* penyakit hati (*liver*) yang diperoleh dari situs *web* Kaggle. Bahasa pemrograman yang digunakan dalam penelitian ini adalah Python. Hasil penelitian menunjukkan bahwa metode klasifikasi *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* berhasil diimplementasikan untuk mendiagnosis penyakit hati (*liver*). Hasil kinerja ketiga metode tersebut, yaitu metode *Regresi Logistik Biner* memperoleh nilai akurasi sebesar 0.98. Metode *Support Vector Machine* memperoleh nilai akurasi sebesar 0.97. Metode *Random Forest* memperoleh nilai akurasi tertinggi sebesar 0.99.

Kata kunci: *Machine Learning*, *Regresi Logistik Biner*, *Random Forest*, *Support Vector Machine*, Penyakit Hati (*Liver*).

ABSTRACT

PERFORMANCE ANALYSIS OF VARIOUS CLASSIFICATION METHODS FOR LIVER DISEASE DIAGNOSIS

By

SALSABILLA JULIA FARHANA

The liver is an organ in the human body that has an important role in the human body. Liver disease is a disorder of the liver or liver that causes the organ to stop functioning properly. The purpose of this research is to apply various classification methods of Binary Logistic Regression, Random Forest, and Support Vector Machine to diagnose liver disease. The application of this classification method is expected to provide information in diagnosing liver disease. This research uses a liver disease dataset sourced from the Kaggle website. The programming language used in this research is Python. The research results indicate that the Binary Logistic Regression, Random Forest, and Support Vector Machine classification methods have been successfully implemented for diagnosing liver disease. The performance results of these three methods are as follows Binary Logistic Regression achieved an accuracy score of 0.98, Support Vector Machine achieved an accuracy score of 0.97, and Random Forest obtained the highest accuracy score of 0.99.

Key words: Machine Learning, Binary Logistic Regression, Random Forest, Support Vector Machine, Liver Disease.

**ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK
DIAGNOSIS PENYAKIT HATI (*LIVER*)**

Oleh

SALSABILLA JULIA FARHANA

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **ANALISIS KINERJA BERBAGAI METODE
KLASIFIKASI UNTUK DIAGNOSIS
PENYAKIT HATI (LIVER)**

Nama Mahasiswa : **Salsabilla Julia Farhana**

Nomor Pokok Mahasiswa : 1917051004

Program Studi : S1 Ilmu Komputer

Jurusan : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing



Prof. Admi Syarif, Ph.D.
NIP. 19670103 199203 1 003



Dewi Asiah Shofiana, S.Komp., M.Kom.
NIP. 19950929 202012 2 030

2. Ketua Jurusan Ilmu Komputer

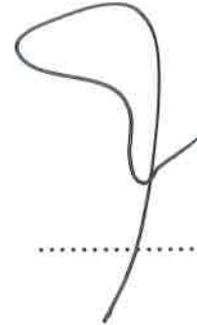


Didik Kurniawan, S.Si., M.T.
NIP. 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

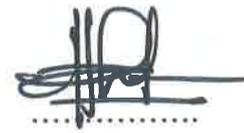
Ketua : Prof. Admi Syarif, Ph.D.



Sekretaris : Dewi Asiah Shofiana, S.Komp., M.Kom.



Penguji Bukan Pembimbing : Dr. rer. nat. Akmal Junaidi, M.Sc.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: 31 Juli 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama: Salsabilla Julia Farhana

NPM: 1917051004

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Analisis Kinerja Berbagai Metode Klasifikasi Untuk Diagnosis Penyakit Hati (*Liver*)” merupakan karya saya sendiri bukan karya orang lain. Semua tulisan yang tertulis dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti bahwa skripsi saya merupakan hasil penjiplakan karya orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya peroleh.

Bandar Lampung, 31 Juli 2023



Salsabilla Julia Farhana

NPM. 1917051004

RIWAYAT HIDUP



Penulis dilahirkan di Bandar Lampung pada tanggal 04 Januari 2001, sebagai anak pertama dari dua bersaudara, dari bapak Herwansyah Syarif dan Ibu Marlenavia. Penulis menyelesaikan pendidikan formal di Taman Kanak-kanak Mathaa'ul Anwar Panjang yang selesai pada tahun 2007, melanjutkan pendidikan dasar di SDN 1 Karang Maritim dan selesai pada tahun 2013, kemudian melanjutkan pendidikan menengah pertama di SMP Negeri 4 Bandar Lampung yang selesai pada tahun 2016, lalu menyelesaikan pendidikan menengah atas di SMA Negeri 10 Bandar Lampung pada tahun 2019.

Pada tahun 2019, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMPTN. Selama menjadi mahasiswa, penulis mengikuti beberapa kegiatan sebagai berikut.

1. Menjadi Asisten Dosen Jurusan Ilmu Komputer mata kuliah Matematika pada tahun 2020 dan mata kuliah Sistem Operasi pada tahun 2021.
2. Menjadi anggota Biro Kesekretariatan Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2019/2020.
3. Menjadi anggota Bidang Keilmuan Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2020/2021.
4. Menjadi Koordinator Divisi Photography pada Pekan Raya Jurusan (PRJ) Ilmu Komputer pada tahun 2021.
5. Menjadi Panitia Penerimaan Mahasiswa baru Jurusan Ilmu Komputer (PRINTER) pada divisi Kakak Asuh pada tahun 2021.

6. Melaksanakan Karya Wisata Ilmiah (KWI) di Desa Tambah Dadi, Kecamatan Purbolinggo, Kabupaten Lampung Timur pada bulan Desember tahun 2019.
7. Melaksanakan Kerja Praktik di PT Jasa Raharja (Persero) Cabang Lampung pada periode 1 tahun 2022.
8. Melaksanakan KKN di Desa Pinang Jaya, Kecamatan Kemiling, Bandar Lampung pada periode 2 tahun 2022.
9. Mengikuti program Kredensial Mikro Mahasiswa Indonesia (KMMI) pada tahun 2021.

MOTTO

“Sesungguhnya sesudah kesulitan itu ada kemudahan, Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain.”

(QS. Al – Insyiroh: 6-7)

“Semangatlah dalam hal yang bermanfaat untukmu, minta tolonglah pada Allah dan jangan malah (patah semangat).”

(HR. Muslim no. 2664)

“Sukses berjalan dari satu kegagalan ke kegagalan yang lain, tanpa kita kehilangan semangat.”

(Abraham Lincoln)

“Kamu tidak harus menjadi hebat untuk memulai, tetapi kamu harus mulai untuk menjadi hebat.”

(Zig Ziglar)

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur saya ucapkan kepada Allah Subhanahu Wa Ta'ala atas segala Rahmat dan Karunia-Nya serta shalawat dan salam senantiasa juga tercurahkan kepada Nabi Muhammad SAW yang memberikan saya kekuatan, kesabaran, serta kelancaran sehingga saya dapat menyelesaikan skripsi ini.

Kupersembahkan karya ini kepada:

Ayah dan Ibuku tercinta atas segala do'a, pengorbanan, perjuangan, kasih sayang, perhatian, dukungan yang selalu kalian berikan kepadaku. Terima kasih karena telah mendidik dan membesarkanku dengan penuh kasih sayang yang tak akan terbalaskan. Kuucapkan juga terima kasih kepada adikku atas dukungan dan do'a yang diberikan kepadaku.

Keluarga Besar Ilmu Komputer 2019

Yang senantiasa memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer.

SANWACANA

Puji syukur kehadiran Allah SWT atas limpahan nikmat, rahmat dan karunia-Nya, shalawat serta salam semoga senantiasa tercurahkan kepada junjungan kita Nabi Muhammad SAW, sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan tepat waktu. Skripsi ini disusun sebagai syarat untuk memperoleh gelar Sarjana Komputer di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dengan judul “Analisis Kinerja Berbagai Metode Klasifikasi Untuk Diagnosis Penyakit Hati (*Liver*)”.

Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun skripsi ini, antara lain:

1. Allah SWT yang telah memberikan limpahan nikmat, rahmat, hidayah, dan karunia-Nya.
2. Kedua orang tua saya, Ayah dan Ibu, juga adik saya yang selalu memberikan doa, dukungan, semangat, motivasi, dan kasih sayang luar biasa tak terhingga. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan di akhirat, Aamiin.
3. Bapak Prof. Admi Syarif, Ph.D. dan ibu Dewi Asiah Shofiana, S.Komp., M.Kom., selaku dosen pembimbing yang selalu sabar membimbing, memberi arahan, dan nasihat, sehingga skripsi ini dapat diselesaikan tepat waktu.
4. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc selaku dosen pembahas sekaligus sekretaris Jurusan Ilmu Komputer Universitas Lampung yang telah memberikan masukan dan saran yang bermanfaat guna menyempurnakan penulisan skripsi ini.
5. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D. selaku pembimbing akademik.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan FMIPA Universitas Lampung.

7. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan banyak ilmu dan pengalaman selama penulis menjalani perkuliahan.
9. Olivia Desti Riana, Sendy Hani Pramita, Devi Rahmadia Fitri, Finka Marisa Geananda Sufie, Zahara Liza Mulyani, Hani Cita Lestari, dan Vira Verina, sahabat-sahabat saya di kampus yang sejak awal kuliah selalu ada dan saling membantu satu sama lain, tempatku menuangkan segala kebahagiaan dan kesedihan. Terima kasih sudah mau menjadi sahabat yang baik.
10. Keluarga Ilmu Komputer 2019 yang tidak bisa disebut satu persatu yang telah bersedia menjadi rekan kelompok, rekan diskusi, dan rekan bercanda. Terima kasih sudah memberi warna dan pengalaman selama masa perkuliahan.

Bandar Lampung, 31 Juli 2023

Salsabilla Julia Farhana

DAFTAR ISI

Halaman

DAFTAR TABEL	iii
DAFTAR GAMBAR	iv
I. PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	5
II. TINJAUAN PUSTAKA	6
2.1. Penelitian Terdahulu.....	6
2.2. Hati (<i>Liver</i>)	13
2.3. Penyakit Hati (<i>Liver</i>)	14
2.4. <i>Data Mining</i>	16
2.5. Klasifikasi.....	18
2.6. <i>Regresi Logistik</i>	20
2.6.1. Kelebihan Metode <i>Regresi Logistik</i>	21
2.6.2. Kekurangan Metode <i>Regresi Logistik</i>	21
2.7. <i>Random Forest</i>	21
2.7.1. Kelebihan Metode <i>Random Forest</i>	23
2.7.2. Kekurangan Metode <i>Random Forest</i>	23
2.8. <i>Support Vector Machine (SVM)</i>	23
2.8.1. Kelebihan Metode <i>Support Vector Machine (SVM)</i>	25
2.8.2. Kekurangan Metode <i>Support Vector Machine (SVM)</i>	25
2.9. Evaluasi Kinerja Klasifikasi	25
III. METODOLOGI PENELITIAN	28
3.1. Waktu dan Tempat Penelitian	28
3.2. Data dan Perangkat Penelitian.....	29
3.2.1. Data.....	29

3.2.2. Perangkat Penelitian	30
3.3. Alur Kerja Penelitian.....	33
3.3.1. Studi Literatur	34
3.3.2. Pengambilan Data.....	34
3.3.3. <i>Preprocessing</i>	34
3.3.4. Pembagian Data.....	35
3.3.5. Pemodelan.....	35
3.3.6. Evaluasi.....	35
3.3.7. Hasil.....	36
IV. HASIL DAN PEMBAHASAN	37
4.1. Hasil Penelitian.....	37
4.1.1. <i>Preprocessing</i>	37
4.1.2. Pemodelan.....	70
4.2. Pembahasan	76
V. SIMPULAN DAN SARAN.....	80
5.1. Simpulan.....	80
5.2. Saran.....	81
DAFTAR PUSTAKA	82

DAFTAR TABEL

Tabel	Halaman
1. Penelitian terdahulu	6
2. <i>Confusion matrix</i>	26
3. Alur waktu penelitian.....	28
4. Atribut <i>dataset</i>	29
5. Transformasi atribut.....	38
6. <i>Missing value</i> pada <i>dataset</i> penyakit hati (<i>liver</i>)	69
7. Iterasi <i>k-fold cross-validation</i> pada <i>Regresi Logistik Biner</i>	71
8. <i>Confusion matrix Regresi Logistik Biner</i>	71
9. Hasil kinerja metode <i>Regresi Logistik Biner</i>	72
10. Iterasi <i>k-fold cross-validation</i> pada <i>Random Forest</i>	73
11. <i>Confusion matrix Random Forest</i>	73
12. Hasil kinerja metode <i>Random Forest</i>	74
13. Iterasi <i>k-fold cross-validation</i> pada SVM.....	75
14. <i>Confusion matrix SVM</i>	75
15. Hasil kinerja metode SVM.....	76
16. Perbandingan kinerja <i>Regresi Logistik Biner</i> , <i>Random Forest</i> , SVM.....	76
17. Perbandingan hasil akurasi.....	77

DAFTAR GAMBAR

Gambar	Halaman
1. SVM Linear	24
2. SVM Non-linear.....	24
3. Alur Kerja Penelitian.	33
4. <i>Plot</i> Distribusi Fitur Numerik	39
5. <i>Heatmap Dataset</i> Penyakit Hati (<i>Liver</i>).....	40
6. Grafik <i>Bar</i> Atribut <i>Gender</i>	41
7. <i>Plot Violin</i> Atribut <i>Age</i>	42
8. <i>Plot Kernel Density Estimate</i> (KDE) Atribut <i>Age</i>	43
9. <i>Plot Violin</i> Atribut Total <i>Bilirubin</i>	44
10. <i>Plot Kernel Density Estimate</i> (KDE) Atribut Total <i>Bilirubin</i>	45
11. <i>Plot Violin</i> Atribut <i>Direct Bilirubin</i>	46
12. <i>Plot Kernel Density Estimate</i> (KDE) Atribut <i>Direct Bilirubin</i>	47
13. <i>Plot Violin</i> Atribut <i>Alkaline Phosphatase</i>	48
14. <i>Plot KDE</i> Atribut <i>Alkaline Phosphatase</i>	49
15. <i>Plot Violin</i> Atribut <i>Alamine Aminotransferase</i>	50
16. <i>Plot Kernel Density Estimate</i> (KDE) <i>Alamine Aminotransferase</i>	51
17. <i>Plot Violin</i> Atribut <i>Aspartate Aminotransferase</i>	52
18. <i>Plot Kernel Density Estimate</i> (KDE) <i>Aspartate Aminotransferase</i>	53
19. <i>Plot Violin</i> Atribut Total <i>Proteins</i>	54
20. <i>Plot Kernel Density Estimate</i> (KDE) Atribut Total <i>Proteins</i>	55
21. <i>Plot Violin</i> Atribut <i>Albumin</i>	56
22. <i>Plot Kernel Density Estimate</i> (KDE) Atribut <i>Albumin</i>	57
23. <i>Plot Violin</i> Atribut <i>Ratio Albumin dan Globulin</i>	58
24. <i>Plot Kernel Density Estimate</i> (KDE) <i>Ratio Albumin dan Globulin</i>	59
25. <i>Scatter Plot</i> Atribut <i>Age</i> dan Atribut <i>Alkaline Phosphatase</i>	60
26. <i>Scatter Plot</i> <i>Alamine</i> dan <i>Aspartate Aminotransferase</i>	61
27. <i>Scatter Plot</i> Atribut <i>Albumin</i> dan Atribut Total <i>Proteins</i>	62

28. <i>Scatter Plot Total Proteins dan Ratio Albumin and Globulin</i>	63
29. <i>Scatter Plot Atribut Ratio Albumin and Globulin dan Albumin</i>	64
30. <i>Bar Chart Atribut Age, Gender, Total Bilirubin, dan Direct Bilirubin</i>	66
31. <i>Bar Chart Atribut Alkaline Phosphatase, Alamine Aminotransferase, dan Aspartate Aminotransferase</i>	68
32. <i>Bar Chart Atribut Total Proteins, Albumin, dan Albumin Globulin Ratio</i>	69

I. PENDAHULUAN

1.1. Latar Belakang

Hati (*Liver*) adalah salah satu organ dalam tubuh manusia yang mempunyai peran penting pada tubuh manusia karena kemampuannya dalam melakukan regenerasi yaitu kemampuan untuk mengganti sel-sel yang rusak (Patimah et al., 2021). Hati (*Liver*) memiliki beberapa fungsi yang krusial, sebagai penetralisir atau penawar racun, pusat metabolisme, mengatur sirkulasi hormon, serta mengatur komposisi darah yang mengandung lemak, protein, dan zat yang lain (Prabiantissa, 2021). Karena fungsi yang beragam tersebut, tidak menjamin hati (*liver*) terus sehat. Jika organ hati (*liver*) mengalami kerusakan maka dapat membahayakan kesehatan yang bahkan dapat menyebabkan kematian (Restiani, 2018). Penyakit hati (*liver*) adalah gangguan pada hati atau *liver* yang menyebabkan organ tersebut berhenti berfungsi dengan baik. Penyakit hati (*liver*) dapat disebabkan oleh banyak faktor, antara lain infeksi virus, kecanduan alkohol, dan penumpukan lemak di hati (Pittara, 2022).

Permasalahan yang sering terjadi saat ini adalah sulitnya mengenali penyakit hati (*liver*) sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Padahal mengetahui gejala penyakit hati (*liver*) sejak dini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan cepat dan tepat. Dengan adanya diagnosis penyakit hati (*liver*) lebih awal dapat meningkatkan tingkat kelangsungan hidup pasien (Pusporani et al., 2019). Diagnosis penyakit hati (*liver*) dapat dilakukan dengan menggunakan berbagai metode klasifikasi.

Salah satu metode statistika yang dapat melakukan pengkategorian adalah klasifikasi. Klasifikasi merupakan suatu proses menemukan model atau fungsi yang dapat membedakan dan menggambarkan kelas pada suatu data atau konsep. Klasifikasi merupakan teknik *data mining* yang dapat digunakan untuk memprediksi keanggotaan kelompok untuk *data instance* (Wijaya, 2018). Terdapat beberapa metode klasifikasi yang dapat digunakan. Dengan melakukan analisis kinerja berbagai metode klasifikasi terhadap gejala-gejala dan data kesehatan tubuh kita dapat mengetahui diagnosis penyakit hati (*liver*) lebih awal. Analisis data tersebut dapat dilakukan untuk diagnosis penyakit hati (*liver*) dengan menggunakan teknik pembelajaran mesin yang efisien.

Beberapa penelitian terdahulu yang menerapkan metode klasifikasi untuk diagnosis penyakit hati (*liver*), diantaranya penelitian yang dilakukan oleh (Patimah et al., 2021) dengan judul “Klasifikasi Penyakit *Liver* Dengan Menggunakan Metode *Decision Tree*”. Data yang digunakan pada penelitian ini adalah *dataset* dari situs Kaggle yang terdiri dari 583 data pasien *liver* dengan 11 atribut. Hasil dari penelitian ini yaitu didapatkan nilai akurasi sebesar 73.33%.

Selanjutnya (Gokiladevi et al., 2022) melakukan penelitian klasifikasi penyakit Ginjal Kronis dengan judul “*Machine Learning Algorithm Selection For Chronic Kidney Disease Diagnosis and Classification*” menggunakan beberapa metode klasifikasi, yaitu SVM, *Random Forest*, *Regresi Logistik*, K-NN, *Decision Tree*. Data yang digunakan adalah *dataset* dari *UCI repository* yang terdiri dari 400 data dengan 24 fitur dan 2 class. Hasil penelitian ini didapatkan bahwa algoritme *Random Forest* lebih baik daripada algoritme lainnya yaitu menghasilkan nilai akurasi terbaik sebesar 98.43%.

Selanjutnya penelitian klasifikasi penyakit hati (*liver*) yang dilakukan oleh (Islam et al., 2018) dengan judul “*Applications of Machine Learning in Fatty*

Live Disease Prediction” menggunakan beberapa metode klasifikasi, yaitu *Random Forest*, SVM, ANN, dan *Regresi Logistik*. Data yang digunakan adalah *dataset* dari *Taipei Medical University Hospital* yang terdiri dari 994 pasien. Hasil penelitian ini didapatkan bahwa algoritme *Regresi Logistik* menunjukkan kinerja yang lebih baik dibandingkan algoritme lainnya yaitu dengan nilai akurasi sebesar 70.7%.

Pada penelitian ini memutuskan untuk menggunakan metode klasifikasi *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine (SVM)* dikarenakan pada penelitian terdahulu memberikan saran untuk penelitian selanjutnya dapat menggunakan metode lain seperti metode *Regresi*, *Random Forest*, SVM. Tujuannya agar mendapatkan nilai akurasi lainnya, sehingga dapat mengetahui atau membandingkan metode mana yang lebih baik untuk klasifikasi *dataset* penyakit hati (*liver*).

Metode *Regresi Logistik* adalah suatu metode analisis statistik yang digunakan untuk menggambarkan hubungan sebab akibat antar variabel bebas (*independen*) dan variabel respon (*dependen*) (Tampil et al., 2017). Model *regresi logistik* yang variabel dependennya terdiri dari dua kategori disebut model *regresi logistik biner* (dikotomi). *Regresi logistik biner* digunakan untuk menganalisis hubungan antara variabel respon (Y) yang berskala dikotomi dan satu atau lebih variabel bebas (X) (Suhendra et al., 2020). Regresi ini digunakan untuk memprediksi kemungkinan semua jenis hasil “ya” atau “tidak” yang biasanya diberi angka 0 atau 1 (Tampil et al., 2017).

Metode selanjutnya adalah *Random Forest*, yaitu algoritme pada *machine learning* yang merupakan kumpulan dari *decision tree* atau pohon keputusan untuk memprediksi suatu kejadian (Basar et al., 2022). *Random Forest* adalah metode yang digunakan untuk klasifikasi dengan membuat beberapa pohon klasifikasi. Metode ini dapat meningkatkan keakuratan hasil dengan membuat

simpul anak untuk setiap *node* (simpul di atasnya) dan membuat pilihan acak (Sandag, 2020).

Metode selanjutnya adalah *Support Vector Machine (SVM)* adalah suatu sistem pembelajaran dengan ruang hipotesis berdasarkan fungsi linier pada sebuah fitur berdimensi tinggi dan dilatih dengan algoritme pembelajaran berdasarkan teori optimasi (Susilowati et al., 2015). Metode SVM dibagi menjadi dua yaitu Linear SVM dan SVM non-linear.

Berdasarkan pembahasan tersebut, penelitian ini akan membahas “**Analisis Kinerja Berbagai Metode Klasifikasi Untuk Diagnosis Penyakit Hati (*Liver*)**”. Tujuan penelitian ini adalah mengklasifikasi diagnosis penyakit hati (*liver*) dengan berbagai metode klasifikasi sehingga masyarakat dapat mendeteksi penyakit hati (*liver*) lebih awal. Data yang digunakan pada penelitian ini adalah *dataset* penyakit hati (*liver*) yang bersumber dari situs web Kaggle.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine (SVM)* untuk mendiagnosis penyakit hati (*liver*).
2. Bagaimana hasil analisis kinerja metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine (SVM)* untuk diagnosis penyakit hati (*liver*).

1.3. Batasan Masalah

Agar permasalahan yang dirumuskan tidak meluas, maka batasan-batasan masalah yang ditentukan pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah *dataset* penyakit hati (*liver*) yang bersumber dari situs web Kaggle.
2. Metode yang digunakan pada klasifikasi prediksi penyakit hati (*liver*) yaitu metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM).
3. Bahasa pemrograman yang digunakan pada penelitian ini adalah *Python*.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM) untuk mendiagnosis penyakit hati (*liver*).
2. Mengetahui hasil kinerja berbagai metode klasifikasi untuk mendiagnosis penyakit hati (*liver*).

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui hasil kinerja klasifikasi dari metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM) untuk mendiagnosis penyakit hati (*liver*).
2. Dapat memberikan informasi dalam mendiagnosis penyakit hati (*liver*) sehingga masyarakat dapat mendeteksi penyakit hati (*liver*) lebih awal.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini dikembangkan berdasarkan beberapa referensi penelitian sebelumnya. Hasil dari beberapa penelitian sebelumnya disajikan pada Tabel 1 berikut.

Tabel 1. Penelitian terdahulu

Penelitian	Data	Metode	Hasil
(Patimah et al., 2021)	Nama data: Penyakit Hati (<i>liver</i>) Jumlah data: 583 data pasien Atribut: 11 Sumber: Kaggle	<i>Decision Tree</i>	Hasil klasifikasi dengan metode <i>Decision Tree</i> accuracy: 73%.
(Singaravelu et al., 2018)	Nama data: Penyakit Hati (<i>liver</i>)	<i>Random Forest, Naïve Bayes, K-star</i>	Metode <i>Random Forest</i> accuracy: 74.2%.

Penelitian	Data	Metode	Hasil
	Jumlah data: 583 data pasien Atribut: 11 Sumber: Kaggle		Metode <i>Naïve Bayes</i> accuracy: 60.6%. Metode K-star accuracy: 67.2%.
(Priya et al., 2018)	Nama data: Penyakit Hati (<i>liver</i>) Jumlah data: 583 data pasien Atribut: 11 Sumber: Kaggle	<i>Support Vector Machine, Bayesian Network, Random Forest, Multilayer Perceptron</i>	Metode <i>Support Vector Machine</i> accuracy: 73.44%. Metode <i>Bayesian Network</i> accuracy: 90.33%. Metode <i>Random Forest</i> accuracy: 80.22% Metode <i>Multilayer Perceptron</i> accuracy: 77.54%.
(Hikmah & Yasa, 2021)	Nama data: Penyakit Hati (<i>liver</i>) Jumlah data: 583 data Atribut: 11	<i>Regresi Logistik, Random Forest, SVM, Neural Network, Naïve Bayes, KNN, Decision Tree</i>	Metode <i>Regresi Logistik</i> accuracy: 72%. Metode <i>Random Forest</i> accuracy: 70%.

Penelitian	Data	Metode	Hasil
	Sumber: UCI <i>Machine Learning repository</i>		Metode SVM <i>accuracy: 65%.</i> Metode <i>Neural Network accuracy: 71%.</i> Metode <i>Naïve Bayes accuracy: 66%.</i> Metode KNN <i>accuracy: 66%.</i> Metode <i>Decision Tree accuracy: 66%.</i>
(Setiawati et al., 2019)	Nama data: Penyakit Hati <i>(liver)</i> Jumlah data: 583 data pasien Atribut: 11 Sumber: Kaggle	<i>Decision Tree</i>	Hasil klasifikasi dengan metode <i>Decision Tree accuracy: 72.67%.</i>
(Yogi et al., 2019)	Nama data: Penyakit Hati <i>(liver)</i>	<i>Random Forest</i>	Hasil klasifikasi dengan metode

Penelitian	Data	Metode	Hasil
	Jumlah data: 303 data pasien Sumber: UCI <i>Machine Learning Repository</i>		<i>Random Forest accuracy: 92.25%.</i>
(Islam et al., 2018)	Nama data: Penyakit Hati (<i>liver</i>) Jumlah data: 994 data pasien Sumber: <i>Taipei Medical University Hospital</i>	<i>Random Forest, Naïve Bayes, Support Vector Machine, Artificial Neural Network, Regresi Logistik</i>	Metode <i>Random Forest accuracy: 6.58%.</i> Metode <i>Support Vector Machine accuracy: 69%.</i> Metode <i>Artificial Neural Network accuracy: 69.1%.</i> Metode <i>Regresi Logistik accuracy: 70.7%.</i>
(Gokiladevi et al., 2022)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 data	<i>Support Vector Machine, Random Forest, Regresi Logistik, K-NN, Decision Tree</i>	Metode <i>Support Vector Machine accuracy: 80%</i> Metode <i>Random Forest accuracy: 98.43%</i>

Penelitian	Data	Metode	Hasil
	Atribut: 24 dan 2 <i>class</i> Sumber: <i>UCI repository</i>		Metode <i>Regresi Logistik</i> accuracy: 96.25% Metode K-NN accuracy: 80.62% Metode <i>Decision Tree</i> accuracy: 97.81%
(Prabiantissa, 2021)	Nama data: <i>Dataset Heart Disease</i> Jumlah data: 303 data pasien Atribut: 14 Sumber: <i>UCI Machine Learning Repository</i>	<i>Support Vector Machine, K-NN, Naïve Bayes</i>	Metode <i>Support Vector Machine</i> accuracy: 82.42%. Metode K-NN accuracy: 63.74%. Metode <i>Naïve Bayes</i> accuracy: 84.62%.
(Harahap et al., 2021)	Nama data: Penyakit Jantung Jumlah data: 296 data pasien Atribut: 14	<i>Random Forest</i>	Hasil klasifikasi dengan metode <i>Random Forest</i> accuracy: 85.3%.

Penelitian	Data	Metode	Hasil
	Sumber: <i>Hungarian Institute of Cardiology</i>		

Beberapa penelitian terdahulu yang menerapkan metode klasifikasi untuk diagnosis penyakit hati (*liver*), diantaranya klasifikasi penyakit hati (*liver*) yang berjudul “Klasifikasi Penyakit *Liver* Dengan Menggunakan Metode *Decision Tree*” yang dilakukan oleh (Patimah et al., 2021). Data yang digunakan pada penelitian ini adalah *dataset* dari situs Kaggle yang terdiri dari 583 data pasien *liver* dengan 11 atribut. Hasil dari penelitian ini yaitu didapatkan nilai akurasi sebesar 73.33%.

Selanjutnya penelitian klasifikasi penyakit hati (*liver*) yang dilakukan oleh (Singaravelu et al., 2018) dengan judul “*Classification of Liver Patient Dataset Using Machine Learning Algorithms*” menggunakan beberapa metode klasifikasi, yaitu *Random Forest*, *Naïve Bayes*, K-star. Data yang digunakan adalah *dataset* dari timur laut Andhra Pradesh, India yang terdiri dari 583 data pasien *liver* dengan 11 atribut. Hasil penelitian ini didapatkan bahwa metode *Random Forest* menunjukkan nilai akurasi tertinggi dibandingkan algoritme lainnya sebesar 74.2%. Sedangkan metode *Naïve Bayes* memberikan nilai akurasi terendah sebesar 60.6%.

Penelitian klasifikasi penyakit hati (*liver*) yang dilakukan oleh (Priya et al., 2018) dengan judul “*Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms*” menggunakan beberapa metode klasifikasi, yaitu SVM, *Bayesnet*, dan MLP. Data yang digunakan adalah *dataset* ILPD (*Indian Liver Patient Dataset*) yang terdiri dari 583 *record* dengan 11 atribut. Hasil dari penelitian ini, yaitu didapatkan hasil bahwa metode *Bayesnet* memberikan hasil nilai akurasi yang lebih baik dibandingkan metode lainnya sebesar 90.33%.

Selanjutnya (Hikmah & Yasa, 2021) melakukan penelitian klasifikasi penyakit hati (*liver*) dengan judul “Perbandingan Hasil Prediksi Diagnosis pada *Indian Liver*

Patient Dataset (ILPD) dengan *Teknik Supervised Learning Menggunakan Software Orange*” menggunakan beberapa metode klasifikasi, yaitu *Regresi Logistik, Random Forest, SVM, Neural Network, Naïve Bayes, KNN, Decision Tree*. Data yang digunakan adalah *dataset* dari *UCI machine learning repository* yang terdiri dari 583 data dengan 11 atribut. Hasil penelitian ini didapatkan bahwa algoritme *Regresi Logistik* lebih baik daripada algoritme lainnya yaitu menghasilkan nilai akurasi terbaik sebesar 72%.

Penelitian klasifikasi penyakit hati (*liver*) dengan judul “*Implementasi Decision Tree Untuk Mendiagnosis Penyakit Liver*” yang dilakukan oleh (Setiawati et al., 2019). *Dataset* yang digunakan pada penelitian ini diambil dari database *UCI Machine Learning Repository*, yaitu *dataset ILPD (Indian Liver Patient Dataset)* yang berisi 583 data klinis dengan 10 atribut. Hasil dari penelitian ini yaitu didapatkan nilai akurasi sebesar 72.67% pada implementasi metode *Decision Tree* untuk mendiagnosis penyakit hati (*liver*).

Selanjutnya penelitian klasifikasi penyakit Jantung yang dilakukan oleh (Yogi et al., 2019) dengan judul “*Expert System for Early Diagnosis of Heart Disease Using Random Forest Method*” menggunakan metode *Random Forest*. Data yang digunakan adalah *dataset UCI Machine Learning Repository* yang terdiri dari 303 data. Hasil dari penelitian ini yaitu didapatkan nilai akurasi sebesar 92.25%.

Penelitian klasifikasi penyakit hati (*liver*) yang dilakukan oleh (Islam et al., 2018) dengan judul “*Applications of Machine Learning in Fatty Liver Disease Prediction*” menggunakan beberapa metode klasifikasi, yaitu *Random Forest, SVM, ANN, dan Regresi Logistik*. Data yang digunakan adalah *dataset* dari *Taipei Medical University Hospital* yang terdiri dari 994 pasien. Hasil penelitian ini didapatkan bahwa algoritme *Regresi Logistik* menunjukkan kinerja yang lebih baik dibandingkan algoritme lainnya yaitu dengan nilai akurasi sebesar 70.7%.

Selanjutnya (Gokiladevi et al., 2022) melakukan penelitian klasifikasi penyakit Ginjal Kronis dengan judul “*Machine Learning Algorithm Selection For Chronic*

Kidney Disease Diagnosis and Classification” menggunakan beberapa metode klasifikasi, yaitu SVM, *Random Forest*, *Regresi Logistik*, K-NN, *Decision Tree*. Data yang digunakan adalah *dataset* dari *UCI repository* yang terdiri dari 400 data dengan 24 fitur dan 2 *class*. Hasil penelitian ini didapatkan bahwa algoritme *Random Forest* lebih baik daripada algoritme lainnya yaitu menghasilkan nilai akurasi terbaik sebesar 98.43%.

Penelitian klasifikasi pada *dataset* penyakit hati (*liver*) yang dilakukan oleh (Prabiantissa, 2021) dengan judul “Klasifikasi Pada *Dataset* Penyakit Hati Menggunakan Algoritme *Support Vector Machine*, K-NN, dan *Naïve Bayes*”. Data yang digunakan adalah *dataset heart disease* dari halaman web *UCI Machine Learning Repository* yang terdiri dari 14 atribut dan 303 *instance*. Hasil penelitian ini didapatkan bahwa algoritme *Naïve Bayes* menunjukkan persentase rata-rata 80.97%, K-NN memiliki persentase yang tidak terlalu baik, dan disimpulkan bahwa algoritme SVM mempunyai performa terbaik dari ketiga algoritme dengan hasil persentase 84.62%.

Selanjutnya penelitian klasifikasi penyakit Jantung yang dilakukan oleh (Harahap et al., 2021) dengan judul “Klasifikasi Diagnosis Penyakit Jantung menggunakan Algoritme *Random Forest*” menggunakan metode *Random Forest*. Data yang digunakan adalah *dataset* dari *Hungarian Institute of Cardiology* yang terdiri dari 14 atribut. Hasil dari penelitian ini yaitu didapatkan nilai akurasi sebesar 85.3%.

2.2. Hati (*Liver*)

Hati (*liver*) adalah salah satu organ yang perannya sangat penting atau vital dalam tubuh manusia karena dapat melakukan regenerasi yaitu kemampuan untuk mengganti sel yang rusak (Patimah et al., 2021). Organ yang juga disebut *liver* ini merupakan organ terbesar di dalam tubuh manusia. Hati (*liver*) terletak di rongga perut kanan bagian atas, tepat di bawah rusuk bagian kanan. Organ ini memiliki dua bagian, yaitu bagian kanan dan kiri. Lobus kanan hati merupakan bagian terbesar

yang ukurannya lima sampai enam kali lebih besar daripada lobus kiri. Berikut beberapa fungsi yang dimiliki Hati (*liver*), antara lain (Makarim, 2022):

1. Membersihkan Darah

Hati (*liver*) berfungsi membersihkan atau menyaring darah dari senyawa-senyawa berbahaya yang ada di dalam tubuh. Seperti senyawa yang berasal dari obat-obatan, alkohol, hingga racun.

2. Memproduksi Protein

Organ yang satu ini bertanggung jawab untuk memproduksi protein, seperti albumin yang berfungsi menjaga cairan dalam sistem sirkulasi tubuh.

3. Metabolisme Protein

Hati (*liver*) juga berperan dalam membantu metabolisme protein dengan mengubah amonia menjadi urea yang dikeluarkan bersama urine oleh ginjal.

4. Penyimpanan Nutrisi

Hati (*liver*) berperan penting dalam proses penyimpanan nutrisi tubuh. Misalnya zat besi, vitamin A, B12, D, dan K, serta asam folat.

Hati (*liver*) bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh, dimana hilangnya fungsi-fungsi tersebut dapat menyebabkan kerusakan yang serius pada tubuh (Pusporani et al., 2019).

2.3. Penyakit Hati (*Liver*)

Penyakit Hati (*liver*) adalah istilah yang digunakan untuk setiap gangguan pada hati atau *liver* yang menyebabkan organ ini tidak dapat berfungsi dengan baik (Pittara, 2022). Hati (*liver*) bertanggung jawab untuk fungsi-fungsi kritis dalam tubuh. Hati (*liver*) adalah satu-satunya organ dalam tubuh yang dapat dengan mudah mengganti sel-sel yang rusak, tetapi jika sel-sel itu hilang, maka *liver* tidak mungkin dapat memenuhi kebutuhan tubuh. Jika organ hati (*liver*) mengalami kerusakan maka kesehatan pun akan terganggu, bahkan dapat menyebabkan kematian. Penyebab

penyakit hati (*liver*) sangat beragam (Pusporani et al., 2019). Berikut ini adalah beberapa penyebab penyakit hati (*liver*) antara lain (Azmi, 2022):

1. Infeksi

Salah satu penyebab penyakit hati (*liver*) adalah infeksi parasit atau virus yang menyerang organ hati. Infeksi ini nantinya memicu peradangan, sehingga menghambat fungsi hati.

2. Gangguan sistem imun

Penyakit di mana sistem kekebalan tubuh menyerang bagian-bagian tertentu dari tubuh (autoimun) juga dapat menimbulkan gangguan pada fungsi hati.

3. Gaya hidup

Penyebab penyakit hati (*liver*) juga dapat dipengaruhi oleh gaya hidup, seperti penyalahgunaan alkohol kronis, pola makan tidak sehat, dan penggunaan obat-obatan tertentu.

4. Genetik

Bila anda memiliki satu atau kedua orangtua yang mempunyai gen abnormal perlu berhati-hati. Pasalnya, gen abnormal dapat diturunkan dan menyebabkan berbagai zat menumpuk dalam hati. Akibatnya, kerusakan hati pun terjadi.

Penyakit *liver* sering disebut sebagai pembunuh diam-diam karena kemungkinan tidak timbulnya gejala. Hati sebenarnya tidak memiliki reseptor rasa sakit. Sulitnya mengenali penyakit hati (*liver*) sejak dini menjadi sebuah permasalahan, bahkan ketika penyakit tersebut sudah menyebar. Padahal mengetahui gejala penyakit hati (*liver*) sejak dini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan cepat dan tepat. Dengan adanya diagnosis penyakit hati (*liver*) lebih awal dapat meningkatkan tingkat kelangsungan hidup pasien (Pusporani et al., 2019).

2.4. Data Mining

Data mining adalah suatu proses yang menggunakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstrak data atau pengetahuan (*knowledge*) secara otomatis (Hermawati, 2013). *Data mining* merupakan bidang ilmu penelitian yang dapat mengolah *database* menjadi pengetahuan yang dapat digunakan untuk mendiagnosis penyakit.

Definisi lainnya *data mining* adalah kombinasi teori dan heuristik yang berfokus pada seluruh proses pengetahuan atau penemuan pola, termasuk pembersihan data, pembelajaran, dan memvisualisasikan hasil. *Data mining* sering disebut *knowledge discovery in database* (KDD) yaitu kegiatan yang mengumpulkan dan menggunakan data historis untuk menemukan keturunan, pola, atau hubungan dalam kumpulan data yang besar. Hasil dari *data mining* ini dapat digunakan untuk memperbaiki pengambilan keputusan di masa yang akan datang (Santosa, 2007).

Tujuan dari adanya *data mining* adalah (Suardika, 2019):

1. *Explanatory*, yaitu untuk menjelaskan beberapa kegiatan observasi atau suatu kondisi.
2. *Confirmatory*, yaitu untuk mengkonfirmasi suatu hipotesis yang sudah ada.
3. *Exploratory*, yaitu untuk menganalisis data baru suatu relasi yang janggal.

Data mining memiliki beberapa fungsi yang dapat digunakan. Berikut adalah fungsi *data mining* secara umum (Suardika, 2019):

1. *Classification*

Fungsi dari *Classification* adalah untuk mengklasifikasikan suatu target *class* ke dalam kategori yang dipilih.

2. *Clustering*

Fungsi dari *clustering* adalah untuk mencari pengelompokan atribut ke dalam segmen - segmen berdasarkan kesamaan.

3. *Association*

Fungsi dari *association* adalah untuk mencari hubungan antara atribut atau *item set*, berdasarkan jumlah *item* yang muncul dan *rule association* yang ada.

4. *Regression*

Fungsi dari *regression* hampir mirip dengan klasifikasi. Fungsi dari *regression* adalah bertujuan untuk mencari prediksi dari model yang sudah ada.

5. *Forecasting*

Fungsi dari *forecasting* adalah untuk memprediksi waktu yang akan datang berdasarkan trend yang telah terjadi di masa lalu.

6. *Sequence Analysis*

Fungsi dari *sequence analysis* adalah untuk mencari pola urutan dari rangkaian peristiwa.

7. *Deviation Analysis*

Fungsi dari *deviation analysis* adalah untuk mencari kejadian langka yang sangat berbeda dari kondisi normal (kejadian abnormal).

Machine learning adalah bidang interdisipliner yang luas berdasarkan ilmu komputer, statistik, ilmu kognitif, teknik, teori pengoptimalan, dan banyak disiplin ilmu matematika dan ilmiah. *Machine learning* secara garis besar dibagi menjadi dua pendekatan yaitu *supervised learning* dan *unsupervised learning*.

Supervised learning, yaitu metode pembelajaran dengan adanya pelatihan dan pelatih. Pada metode ini kita mengenal dua istilah yang sangat populer yaitu pelatihan (*training*) dan pengujian (*testing*) (Santosa, 2007). Metode *supervised learning* didasarkan pada kumpulan data yang berlabel. Kumpulan sampel digunakan untuk merangkum karakteristik distribusi ukuran perilaku di setiap aplikasi untuk membentuk model perilaku data. *Supervised learning* selanjutnya dikelompokkan ke dalam masalah klasifikasi dan regresi. Masalah klasifikasi

muncul ketika variabel *output* bersifat kategoris. Di sisi lain masalah regresi muncul ketika variabel *output* adalah nilai riil. *Supervised learning* memiliki beberapa model algoritme, seperti *Random Forest*, *Support Vector Machine*, *Logistic Regression*, *Naive Bayes Classifier*, *Metode Rocchio*, *Back-propagation Linear regression*, *K-Nearest Neighbor*, *Decision Tree*, dan *Neural Network* (Roihan et al., 2020).

Unsupervised learning, yaitu metode tanpa adanya latihan (*training*) dan tanpa adanya seorang guru (*teacher*) dalam melaksanakan tugasnya (*task*). Guru disini adalah label (*output*) (Santosa, 2007). *Unsupervised learning* didasarkan pada pengumpulan data yang tidak berlabel untuk menggambarkan struktur yang tersembunyi. *Unsupervised learning* dikelompokkan dalam masalah *clustering* dan asosiasi. Masalah pengelompokan (*clustering*) adalah tempat untuk menemukan pengelompokan yang melekat dalam data. Masalah asosiasi adalah aturan yang menggambarkan sebagian besar data yang ada. *Unsupervised learning* memiliki beberapa algoritme populer seperti *k-means*, *Aprior*, *Independent Subspace Analysis (ISA)* (Roihan et al., 2020).

2.5. Klasifikasi

Klasifikasi merupakan suatu proses menemukan model atau fungsi yang dapat membedakan dan menggambarkan kelas pada suatu data atau konsep. Klasifikasi merupakan teknik *data mining* yang dapat digunakan untuk memprediksi keanggotaan kelompok untuk *data instance* (Wijaya, 2018). Klasifikasi adalah salah satu komponen dari *data mining* agar data bisa memberikan makna atau bisa dimanfaatkan yang disebut juga dengan *supervised learning* (Suardika, 2019).

Klasifikasi data mempunyai dua langkah pada prosesnya. Pada langkah pertama, sebuah model dibangun yang menggambarkan sekumpulan kategori data atau konsep dari populasi data yang diberikan. Model dibuat dengan menganalisis data pelatihan yang dijelaskan oleh atribut. Setiap *tuple* diasumsikan untuk dimiliki oleh

kelas yang telah ditentukan, seperti ditentukan oleh salah satu atribut, yang dinamakan *class label attribute* (Suardika, 2019).

Setiap teknik klasifikasi menggunakan algoritme pembelajaran untuk mendapatkan model yang paling sesuai dengan hubungan antara himpunan atribut dan label kelas dalam data masukan. Biasanya masukan dari model klasifikasi merupakan sekumpulan *record (training set)*. Setiap *record* meliputi himpunan attributes, salah satunya adalah kelas. Model atau pola atribut kelas merupakan fungsi dari nilai atribut lainnya. *Test set* digunakan untuk mengetahui akurasi model tersebut (Hermanto et al., 2020). Biasanya *dataset* yang akan disampaikan dibagi menjadi *training set* dan *test set*, dimana *training set* digunakan untuk membangun model dan *test set* digunakan untuk validasi (Hermawati, 2013).

Proses klasifikasi didasarkan pada empat komponen antara lain (Iriadi, 2013):

a. Kelas

Variabel *dependen* yang bersifat kategoris yang merepresentasikan ‘*label*’ yang terkandung pada objek.

b. *Predictor*

Variabel *independen* yang direpresentasikan oleh karakteristik (atribut) data.

c. *Training dataset*

Satu set data (*dataset*) yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang sesuai berdasarkan *predictor*.

d. *Testing dataset*

Berisi data baru yang akan diklasifikasikan menurut model yang telah dibuat dan akurasi klasifikasi dievaluasi

2.6. Regresi Logistik

Regresi Logistik adalah suatu metode analisis statistik yang digunakan untuk menggambarkan hubungan sebab akibat antar variabel bebas (*independen*) dan variabel respon (*dependen*) (Tampil et al., 2017). Variabel bebas (*independen*) dapat berupa data kategorik atau kontinu, sedangkan variabel respon (*dependen*) harus dalam skala kategorik (Pusporani et al., 2019). Model *regresi logistik* yang variabel dependennya terdiri dari dua kategori disebut model *regresi logistik biner* (dikotomi). *Regresi logistik biner* digunakan untuk menganalisis hubungan antara variabel respon (Y) yang berskala dikotomi dan satu atau lebih variabel bebas (X). Variabel respon (*dependen*) (Y) adalah data kualitatif dikotomi, dengan nilai 1 menunjukkan adanya sebuah karakteristik dan nilai 0 menunjukkan tidak adanya sebuah karakteristik (Suhendra et al., 2020). Regresi ini digunakan untuk memprediksi kemungkinan semua jenis hasil “ya” atau “tidak” yang biasanya diberi angka 0 atau 1. Model *regresi logistik biner* dipakai apabila variabel responnya membentuk 2 kategori bernilai 0 dan 1, sebagai akibatnya mengikuti distribusi Bernoulli menjadi berikut (Tampil et al., 2017):

Adapun bentuk model regresi logistiknya ditunjukkan pada Persamaan 1 berikut:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})} \dots \dots \dots (1)$$

Adapun keterangan sebagai berikut:

β_0 adalah koefisien konstanta

β_1 adalah koefisien variabel x_{1i}

β_p adalah koefisien variabel x_{pi}

$\pi(x_i)$ adalah peluang terjadinya kejadian yang “sukses” yaitu $y = 1$ dengan nilai probabilitas, $0 \leq \pi(x_i) \leq 1$ dinyatakan sebagai kejadian sukses dan 0 berarti gagal.

Untuk memudahkan estimasi parameter *regresi*, maka transformasikan $\pi(x_i)$ pada persamaan di atas untuk menghasilkan bentuk *regresi logistik*, dapat dihitung dengan rumus Persamaan 2 sebagai berikut:

$$g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \dots \dots \dots (2)$$

$g(x)$ merupakan fungsi hubungan dari model *Regresi Logistik* yang disebut sebagai fungsi hubungan logit.

2.6.1. Kelebihan Metode *Regresi Logistik*

Metode *Regresi Logistik* memiliki beberapa kelebihan sebagai berikut (Kresna, 2019):

1. Tidak ada asumsi normalitas dan heteroskedastisitas untuk variabel *independen* yang digunakan dalam model, sehingga asumsi klasik tidak perlu diuji meskipun terdapat lebih dari satu variabel *independen*.
2. Variabel *independen* dapat berupa campuran variabel kontinu, dikotomis, dan distrik.
3. Tidak memerlukan batasan pada variabel *independen*.
4. Tidak mengharuskan variabel bebasnya dalam bentuk interval.

2.6.2. Kekurangan Metode *Regresi Logistik*

Metode *Regresi Logistik* memiliki kekurangan, yaitu rentang terhadap *underfitting* pada data set yang kelasnya tidak seimbang, sehingga akan menghasilkan akurasi yang rendah (Rianto, 2015).

2.7. *Random Forest*

Random Forest adalah algoritme pada *machine learning* yang merupakan kumpulan dari *decision tree* atau pohon keputusan untuk memprediksi suatu kejadian (Basar et al., 2022). Seperti namanya, “*Forest*” atau “hutan” dibentuk dari banyak *tree* (pohon) yang diperoleh melalui proses *bagging* atau *bootstrap aggregating* (Trivusi, 2022a). Metode *Random Forest* adalah pengembangan dari metode *Classification and Regression Tree* (CART). *Random Forest* adalah metode yang digunakan untuk klasifikasi dengan membuat beberapa pohon klasifikasi. Metode ini dapat meningkatkan keakuratan hasil dengan membuat simpul anak untuk setiap *node* (simpul di atasnya) dan membuat pilihan acak (Sandag, 2020).

Selanjutnya hasil klasifikasi dari setiap pohon diakumulasikan dan dipilih hasil klasifikasi yang paling banyak muncul. Metode ini terdiri dari *root node*, *internal node*, dan *leaf node*. *Root node* merupakan simpul yang berada paling atas atau sering disebut sebagai akar dari pohon keputusan. *Internal node* adalah *node* bercabang, di mana *node* tersebut memiliki setidaknya dua *output* dan hanya satu *input*. *Leaf node* atau *node* terminal adalah *node* terakhir dengan hanya satu *input* dan tidak memiliki *output*. Pohon keputusan dimulai dengan menghitung nilai *entropy* sebagai penentu tingkat ketidakmurnian atribut dan nilai perolehan *information gain*. Untuk menghitung nilai *entropy* menggunakan rumus seperti pada Persamaan 3 sedangkan nilai *information gain* menggunakan Persamaan 4 (Sandag, 2020).

Penghitungan pada metode *Random Forest* dapat dirumuskan seperti pada Persamaan 3 sebagai berikut:

$$Entropy(Y) = - \sum_i p(c|Y) \log_2 p(c|Y) \dots \dots \dots (3)$$

Adapun keterangan sebagai berikut:

Y = Himpunan kasus

P(c|Y) = Proporsi nilai Y terhadap kelas c

Information gain dapat dihitung dengan rumus Persamaan 4 sebagai berikut:

$$Information\ Gain(Y, \alpha) = Entropy(Y) - \sum_{v \in values(\alpha)} \frac{|Y_v|}{|Y_\alpha|} Entropy(Y_v) \dots \dots (4)$$

Adapun keterangan sebagai berikut:

Values (α) = Nilai yang mungkin dalam himpunan kasus α

Y_v = Subkelas dari Y dengan kelas v yang berhubungan dengan kelas α .

Y_α = Semua nilai yang sesuai dengan α .

2.7.1. Kelebihan Metode *Random Forest*

Metode *Random Forest* memiliki beberapa kelebihan sebagai berikut (Trivusi, 2022a):

1. Kuat terhadap data *outlier* (pencilan data).
2. Bekerja dengan baik dengan data non-linear.
3. Risiko *overfitting* lebih rendah.
4. Bekerja secara efisien dengan kumpulan data besar.
5. Akurasi yang lebih baik daripada algoritme klasifikasi lainnya.

2.7.2. Kekurangan Metode *Random Forest*

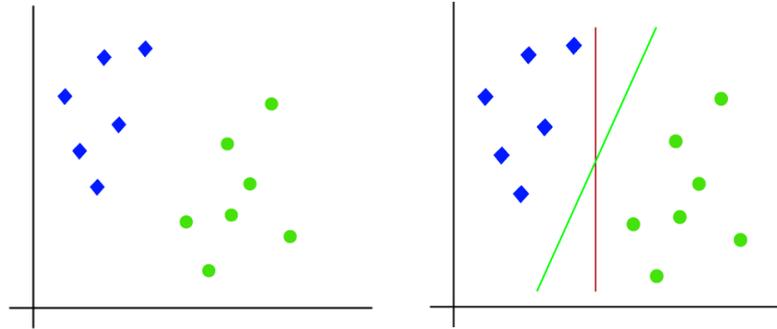
Metode *Random Forest* memiliki beberapa kekurangan sebagai berikut (Trivusi, 2022a):

1. *Random Forest* cenderung bias saat berhadapan dengan variabel kategorikal.
2. Waktu komputasi pada *dataset* berskala besar relatif lambat.
3. Tidak cocok untuk metode linier dengan banyak fitur *sparse*.

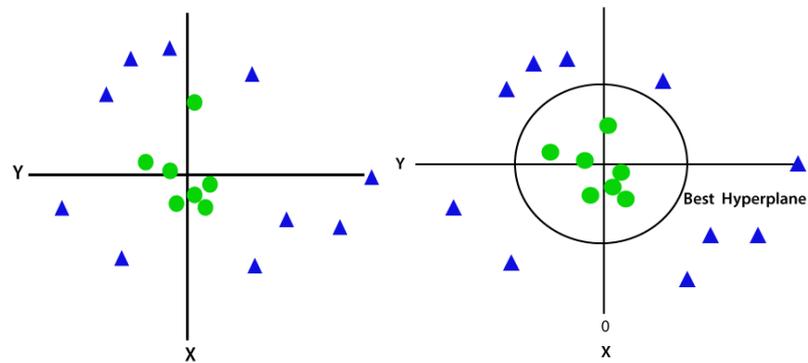
2.8. *Support Vector Machine* (SVM)

Metode *Support Vector Machine* (SVM) pertama kali dikenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian dari konsep-konsep unggulan bidang *pattern recognition*. *Support Vector Machine* (SVM) adalah sistem pembelajaran dengan ruang hipotesis berdasarkan fungsi linier pada sebuah fitur berdimensi tinggi dan dilatih dengan algoritme pembelajaran berdasarkan teori optimasi (Susilowati et al., 2015). Tingkat akurasi dalam model dihasilkan oleh proses peralihan dengan SVM sangat tergantung pada fungsi kernel dan parameter yang digunakan. Berdasarkan karakteristiknya, metode SVM dibagi menjadi dua yaitu Linear SVM dan SVM

non-linear. SVM linier adalah data dipisahkan secara linier dengan memisahkan kelas pada *hyperplane* dengan *soft margin*. SVM Non-linier adalah mengimplementasikan fungsi kernel trick dengan ruang berdimensi tinggi seperti pada Gambar 1 dan 2 berikut (Puspitasari et al., 2018).



Gambar 1. SVM Linear (Trivusi, 2022b)



Gambar 2. SVM Non-linear (Trivusi, 2022b)

Konsep SVM adalah menemukan pemisah yang optimal pada *hyperplane*. *Hyperplane* pemisah terbaik ditemukan dengan mencari nilai $f(x)$ pada *margin hyperplane* (Kurnianto et al., 2018). Menghitung nilai $f(x)$ menggunakan Persamaan 5 sebagai berikut:

$$f(x) = \text{sign} (\sum_{i=1}^n a_i y_i K(x, x_i) + b) \dots \dots \dots (5)$$

SVM non-linear menggunakan fungsi kernel untuk mempermudah perhitungan. Kernel yang umum digunakan adalah RBF (Kurnianto et al., 2018).

2.8.1. Kelebihan Metode *Support Vector Machine* (SVM)

Metode *Support Vector Machine* (SVM) memiliki beberapa kelebihan sebagai berikut (Wibawa et al., 2018):

1. Pengklasifikasi *Support Vector Machine* (SVM) menghasilkan hasil yang sangat akurat dan bekerja dengan baik di ruang berdimensi tinggi.
2. Mudah diimplementasikan.

2.8.2. Kekurangan Metode *Support Vector Machine* (SVM)

Metode *Support Vector Machine* (SVM) memiliki beberapa kekurangan sebagai berikut (Wibawa et al., 2018):

1. Sulit digunakan untuk masalah besar, skala besar yang dimaksud adalah jumlah sampel yang diproses.
2. Jenis kernel metode *Support Vector Machine* (SVM) berpengaruh pada akurasi sistem.

2.9. Evaluasi Kinerja Klasifikasi

Proses akhir klasifikasi penelitian ini yaitu evaluasi kinerja klasifikasi. Evaluasi berfungsi untuk mengetahui nilai performa dari metode yang diterapkan. Pada tahap ini pengujian model penelitian dilakukan dengan metode *Confusion Matrix* yaitu model yang memiliki nilai fungsi untuk melakukan analisis apakah klasifikasi pada metode yang dinilai memiliki label baik atau buruk (Basar et al., 2022).

Confusion matrix merupakan suatu matriks yang menampilkan visualisasi kinerja dari algoritme klasifikasi menggunakan data dalam matriks yang membagi klasifikasi prediksi dalam bentuk *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berikut Tabel 2 *confusion matrix* (Desiani, 2022)

Tabel 2. *Confusion matrix* (Desiani, 2022)

Kelas	Prediksi <i>Positive</i>	Prediksi <i>Negative</i>	Total
Aktual <i>Positive</i>	TP	FP	P
Aktual <i>Negative</i>	FN	TN	N
Total	P'	N'	P+N

Keterangan empat istilah dalam *confusion matrix* yaitu sebagai berikut:

1. *True Positive* (TP): data aktual yang berada di kelas positif dan model memprediksi positif.
2. *True Negative* (TN): data aktual yang berada di kelas negatif dan model memprediksi negatif.
3. *False Positive* (FP): data aktual yang berada di kelas negatif atau netral, namun model memprediksi positif.
4. *False Negative* (FN): data aktual yang berada di kelas positif atau netral, namun model memprediksi negatif.

TP dan TN menunjukkan bahwa *classifier* mengidentifikasi tupel dengan benar, yang berarti bahwa tupel positif dikenali sebagai tupel positif dan tupel negatif sebagai negatif. Sedangkan FP dan FN menyatakan bahwa *classifier* salah dalam mendeteksi tupel, artinya tupel positif terdeteksi sebagai negatif dan tupel negatif sebagai positif. Evaluasi menggunakan *Confusion matrix* akan menghasilkan nilai *Accuracy*, *Recall*, *Precision*, serta *F1-Score* (Wibowo, 2020).

1. *Accuracy*

Accuracy adalah presentasi ketepatan *record* data diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Perhitungan *accuracy* dapat dilihat pada Persamaan 6:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \dots \dots \dots (6)$$

2. *Recall*

Recall digunakan untuk menunjukkan persentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif. Perhitungan *recall* dapat dilihat pada Persamaan 7:

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots \dots \dots (7)$$

3. *Precision*

Precision digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif. Perhitungan *precision* dapat dilihat pada Persamaan 8:

$$Precision = \frac{TP}{TP+FP} \times 100\% \dots \dots \dots (8)$$

4. *F1-Score*

F1-Score adalah rata-rata dari nilai *precision* dan *recall*. Perhitungan *F1-Score* dapat dilihat pada Persamaan 9:

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \dots \dots \dots (9)$$

3.2. Data dan Perangkat Penelitian

Adapun data dan perangkat penelitian yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

3.2.1. Data

Data yang digunakan pada penelitian ini adalah *dataset* penyakit hati (*liver*) yang bersumber dari situs *web* Kaggle yang dikembangkan oleh Vladimir Osin seorang Senior *Machine Learning Engineer* at ASML, Eindhoven, North Brabant, Netherlands yang terdiri dari 583 data dengan 11 atribut dengan format *csv*. *Dataset* penyakit hati (*liver*) yang bersumber dari situs *web* Kaggle yang akan digunakan dapat diakses melalui laman berikut: <https://www.kaggle.com/code/resolut/eda-indian-liver-patient-records>. Adapun atribut-atribut pada *dataset* penyakit hati (*liver*) yang ditampilkan pada Tabel 4 berikut.

Tabel 4. Atribut *dataset*

No	Atribut	Tipe Atribut	Keterangan
1.	<i>Age</i>	<i>Numeric</i>	Umur dari pasien
2.	<i>Gender</i>	<i>Text</i>	Jenis kelamin pasien
3.	TB	<i>Numeric</i>	Total Bilirubin
4.	DB	<i>Numeric</i>	<i>Direct Bilirubin</i>
5.	Alkphos	<i>Numeric</i>	<i>Alkaline Phosphatase</i>
6.	SGPT	<i>Numeric</i>	<i>Alanine Aminotransferase</i>
7.	SGOT	<i>Numeric</i>	<i>Aspartate Aminotransferase</i>
8.	TP	<i>Numeric</i>	<i>Total Proteins</i>
9.	ALB	<i>Numeric</i>	<i>Albumin</i>
10.	A/G Ratio	<i>Numeric</i>	<i>Ratio Albumin dan Globulin Ratio</i>
11.	<i>Class Variabel</i>	<i>Numeric</i>	Terdapat 2 kelas, yaitu kelas 1 dan kelas 2

3.2.2. Perangkat Penelitian

1. Perangkat Keras

- a. *System Manufacture* : *Acer*
- b. *Processor* : *Intel Core I3 3240*
- c. *Installed RAM* : *2.00 GB*
- d. *Gpu* : *Intel® HD Graphics 2500*
- e. *Penyimpanan* : *HDD 500GB*

2. Perangkat Lunak

- a. *Sistem Operasi: Windows 7 Pro 64-Bit.*

- b. *Tools*

- a) *Python Programming*

Python adalah bahasa pemrograman komputer tingkat tinggi, yang berorientasi objek, interpretatif yang ditemukan pada tahun 1989 oleh Guido van Rossum dan menerbitkan edisi publik pertamanya pada tahun 1991 (Chen & Liu, 2022).

- b) *Google Colaboratory*

Google Colaboratory adalah *coding environment* untuk bahasa pemrograman *Python* dalam format "*notebook*" seperti *Jupyter notebook*. *Google Colaboratory* adalah layanan yang mereplikasi *Jupyter notebook* yang berbasis *cloud Google*. *Google Colaboratory* dapat digunakan dengan browser seperti *Opera*, *Mozilla*, *Google Chrome*, dll (Reyvan, 2021).

- c) *Jupyter notebook*

Jupyter merupakan perangkat lunak yang bersifat *open source* dan *servis* dalam komputasi interaktif dalam berbagai bahasa pemrograman. *JupyterLab* adalah *interactive development environment* berbasis *web* untuk *jupyter notebook* kode program, dan data. *JupyterLab* secara fleksibel mendukung *workflow* untuk data *sains*, komputasi ilmiah, dan *machine learning*. *Jupyter notebook* adalah aplikasi *web* yang bersifat

open source yang ditujukan untuk pembersihan dan konversi data, simulasi numerik, visualisasi data, pemodelan statistik, *machine learning* (Retnoningsih, 2020).

d) *Web Browser*

Web browser merupakan perangkat lunak yang digunakan untuk mengambil dan menampilkan informasi dari *server web*. Semua komponen web, termasuk teks, gambar, dan komponen lainnya dapat dilihat di *web browser*. Contoh dari *web browser*, yaitu *Mozilla Firefox*, *Google Chrome*, *Internet Explorer*, dan lain sebagainya (Kuswandi et al., 2021).

e) *Microsoft Excel*

Microsoft Excel adalah program perangkat lunak yang memungkinkan pengguna untuk mengolah dan menghitung data *numerik* (angka). Data diproses menggunakan rumus *spreadsheet*. *Excel* pertama kali terbit pada tahun 1985 dan terus berkembang hingga saat ini (Eugenia, 2021).

c. *Library*

Library atau *package* yang digunakan pada penelitian ini antara lain:

a) *Pandas*

Pandas adalah *library* atau *package* pada *Python* yang menyediakan struktur data yang cepat, fleksibel, dan ekspresif yang berguna untuk melakukan *preprocessing* dan analisis data. *Pandas* bertujuan untuk memudahkan programmer *Python* untuk bekerja dengan data nyata, yang sebagian besar dalam bentuk data yang tersusun dalam tabel. *Library pandas* dapat mengolah data dalam bentuk csv, txt, excel, dan lain sebagainya (Khasanah, 2021).

b) *Numpy*

Numpy (*Numerical Python*) merupakan *library* atau *package* pada *Python* yang berfokus pada komputasi ilmiah yaitu digunakan untuk

menghitung operasi matematika pada *array* atau larik, mencari nilai *max* dan *min*, membuat vektor matriks, dan lain sebagainya. Keunggulan dari *numpy* adalah dapat terintegrasi dengan beberapa bahasa pemrograman lainnya (Khasanah, 2021).

c) *Matplotlib*

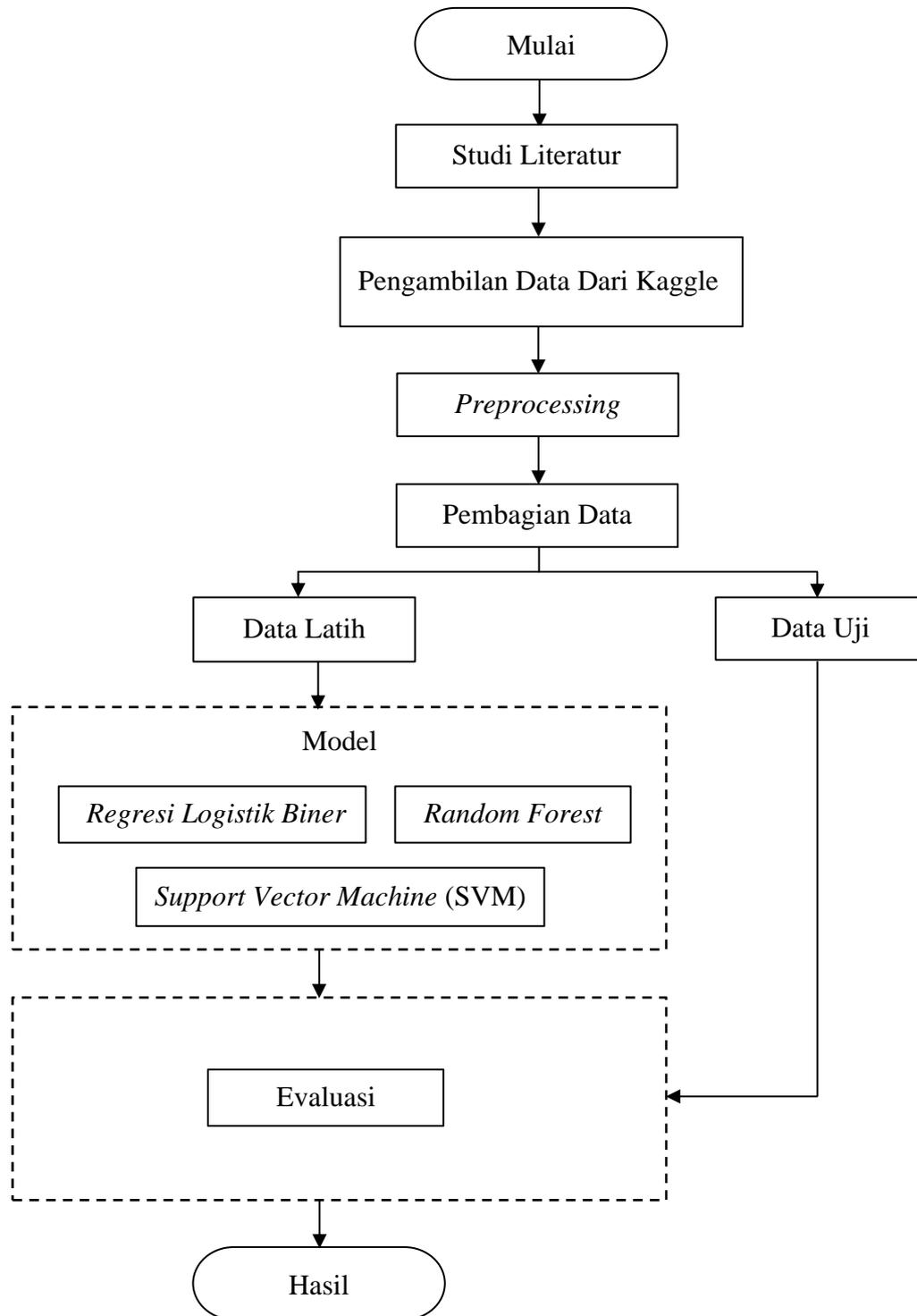
Matplotlib adalah *library Python* yang berfokus pada visualisasi data seperti *plotting*. Visualisasi dari *matplotlib* adalah tampilan grafik dengan satu atau lebih sumbu. *Matplotlib* dapat digunakan pada visualisasi data dalam 2D dan 3D (Khasanah, 2021).

d) *Scikit Learn*

Library *Scikit Learn* merupakan library *Machine Learning* berbasis *Python* yang biasa digunakan dalam *Data Science*. Kelebihan *Scikit Learn* adalah kecepatannya dalam penggunaan API yang mudah dalam melakukan tolak ukur pada *dataset* yang berbeda. Library *Scikit Learn* memudahkan kita dalam proses analisis data yang menggunakan algoritma *machine learning* seperti algoritma *Regresi*, *Naive Bayes*, *Clustering*, *Decision Tree*, *Data Preprocessing Tool*, dan lain sebagainya (Khasanah, 2021).

3.3. Alur Kerja Penelitian

Pada penelitian ini melalui beberapa tahapan pada Gambar 3 berikut.



Gambar 3. Alur Kerja Penelitian.

3.3.1. Studi Literatur

Studi Literatur adalah kegiatan menelusuri dan mengkaji penelitian terdahulu, teori-teori yang mendukung proses penelitian. Langkah ini dimaksudkan sebagai referensi atau acuan untuk memperkuat penelitian yang sedang berlangsung yang ada pada penelitian sebelumnya. Referensi berasal dari berbagai buku, jurnal, dan artikel *website* yang berkaitan dengan tujuan dan masalah penelitian.

3.3.2. Pengambilan Data

Data yang digunakan pada penelitian ini merupakan data sekunder. Pengambilan data pada penelitian ini dilakukan dengan cara mengambil *dataset* penyakit hati (*liver*) yang bersumber dari situs *web* Kaggle yang dikembangkan oleh Vladimir Osin seorang *Senior Machine Learning Engineer at ASML, Eindhoven, North Brabant, Netherlands* terdiri dari 583 data dengan 11 atribut dengan format *Comma Separated Values* (CSV). Dari data tersebut, 416 data merupakan pasien yang terkena penyakit hati (*liver*) dan 167 data pasien yang tidak terkena penyakit hati (*liver*). *Dataset* penyakit hati (*liver*) yang akan digunakan dapat diakses melalui laman berikut: <https://www.kaggle.com/code/resolut/eda-indian-liver-patient-records>.

3.3.3. Preprocessing

Tahap penelitian selanjutnya adalah *preprocessing* merupakan langkah penting dalam melakukan analisis klasifikasi yang bertujuan untuk membersihkan data dari unsur yang tidak berguna antara lain, membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, dan lain sebagainya. Pada tahap ini data mentah yang telah diperoleh diubah ke dalam format yang efisien agar dapat lebih mudah digunakan ke tahap pengolahan data selanjutnya agar data lebih mudah untuk dipahami.

3.3.4. Pembagian Data

Tahapan selanjutnya setelah *preprocessing* yaitu pembagian data, data dibagi menjadi dua yaitu data latih (*training data*) dan data uji. Data latih (*training data*) adalah data yang digunakan untuk melakukan proses learning pada *dataset* penyakit hati (*liver*), sedangkan Data uji (*testing data*) adalah data yang digunakan dalam proses pengujian terhadap hasil learning *dataset* penyakit hati (*liver*). Pada penelitian ini sebanyak 80% data akan menjadi data latih (*training data*) dan 20% data akan menjadi data uji (*testing data*). Tujuan dilakukannya pembagian data adalah untuk menguji model klasifikasi dalam keadaan *dataset* yang berbeda agar terhindar dari *overfitting*, yaitu keadaan dimana model terlalu fokus pada training *dataset* tertentu, sehingga tidak dapat membuat prediksi dengan tepat ketika diberikan kumpulan *dataset* lain.

3.3.5. Pemodelan

Setelah pembagian data, tahap selanjutnya adalah melakukan pemodelan dengan menggunakan dua metode, yaitu metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine (SVM)*. *Regresi logistik biner* digunakan untuk menganalisis hubungan antara variabel respon (Y) yang berskala dikotomi dan satu atau lebih variabel bebas (X). *Random Forest* adalah metode yang digunakan untuk klasifikasi dengan membuat beberapa pohon klasifikasi. *Support Vector Machine (SVM)* adalah suatu sistem pembelajaran dengan ruang hipotesis berdasarkan fungsi linier pada sebuah fitur berdimensi tinggi dan dilatih dengan algoritme pembelajaran berdasarkan teori optimasi. Melihat hasil nilai akurasi dari ketiga metode, metode mana yang lebih baik antara ketiga metode.

3.3.6. Evaluasi

Tahap selanjutnya adalah evaluasi hasil pengklasifikasian diagnosis penyakit hati (*liver*) dengan model *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine (SVM)*. Evaluasi berfungsi untuk mengetahui nilai performa dari metode

yang diterapkan seperti *Accuracy*, *Precision*, *Recall*, serta *F1-Score*. Penelitian ini menggunakan model *Confusion Matrix* merupakan model yang memiliki nilai fungsi untuk menentukan kinerja klasifikasi untuk menilai apakah sudah baik dan tidaknya, atau mengukur tingkat akurasi pada pemodelan klasifikasi yang dapat dilihat dari parameter pengukurannya. Hasil evaluasi ini akan menunjukkan apakah model klasifikasi yang digunakan dapat bekerja dengan baik.

3.3.7. Hasil

Tahapan selanjutnya adalah hasil, langkah ini dilakukan untuk mengkonfirmasi hasil uji akurasi klasifikasi diagnosis penyakit hati (*liver*) dengan metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM). Hasil akurasi dari langkah ini menunjukkan bahwa alat pemodelan yang digunakan dalam proses klasifikasi bekerja dengan baik atau tidak, langkah ini juga menunjukkan interpretasi dari hasil yang diperoleh dengan mengolah semua data untuk mencapai kesimpulan untuk semua objek yang diteliti menggunakan metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM).

V. SIMPULAN DAN SARAN

5.1. Simpulan

Pada penelitian Analisis Kinerja Berbagai Metode Klasifikasi Untuk Diagnosis Penyakit Hati (*Liver*) menggunakan Metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* dapat disimpulkan sebagai berikut.

1. Penelitian ini sudah berhasil mengimplementasikan metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* untuk mendiagnosis penyakit hati (*liver*). Data yang digunakan adalah data penyakit hati (*liver*) yang berasal dari Kaggle dengan jumlah 583 data yang terdiri dari 11 atribut dan proporsi pembagian 80% data *training* dan 20% data *testing*.
2. Hasil kinerja metode *Regresi Logistik Biner*, *Random Forest*, dan *Support Vector Machine* (SVM) untuk diagnosis penyakit hati (*liver*) yaitu metode *Regresi Logistik Biner* memperoleh nilai *accuracy* 0.98, *precision* 0.98, *recall* 0.98, dan *f1-score* 0.98. Metode *Random Forest* memperoleh nilai *accuracy* 0.99, *precision* 0.99, *recall* 0.99, dan *f1-score* 0.99. Metode *Support Vector Machine* memperoleh nilai *accuracy* 0.97, *precision* 0.98, *recall* 0.97, dan *f1-score* 0.97. Berdasarkan hasil tersebut, metode *Random Forest* lebih unggul dibandingkan metode *Regresi Logistik Biner* dan *Support Vector Machine* (SVM) yang menghasilkan nilai akurasi tertinggi, yaitu 0.99 sementara metode *Regresi Logistik Biner* dan *Support Vector Machine* (SVM) juga sudah bekerja dengan baik dalam mendiagnosis penyakit hati (*liver*).

5.2. Saran

Adapun saran yang diberikan pada penelitian ini adalah sebagai berikut.

1. Diharapkan penelitian selanjutnya dapat menggunakan metode klasifikasi lainnya agar dapat memberikan hasil klasifikasi yang lebih baik dalam mendiagnosis penyakit hati (*liver*) dan dapat dilakukan juga perbandingan lainnya.
2. Diharapkan penelitian selanjutnya dapat mengimplementasikan hasil penelitian ke dalam bentuk sistem web atau mobile yang dapat digunakan oleh tenaga kesehatan dalam mendiagnosis penyakit hati (*liver*).
3. Diharapkan penelitian selanjutnya dapat menggunakan *dataset* yang lebih baik dengan variasi atribut yang lebih luas sehingga dapat membantu menganalisis kinerja metode klasifikasi dengan lebih baik.

DAFTAR PUSTAKA

- Azmi, N. (2022). *Penyakit Liver (Penyakit Hati)*. Hellosehat.Com. <https://hellosehat.com/pencernaan/hati/penyakit-liver/>
- Basar, T. F., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest*. 6(3), 1426–1433.
- Chen, X., & Liu, W. (2022). The Value of Python Programming in General Education and Comprehensive Quality Improvement of Medical Students Based on a Retrospective Cohort Study. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/4043992>
- Desiani, A. (2022). Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati. *Simkom*, 7(2), 104–110. <https://doi.org/10.51717/simkom.v7i2.96>
- Eugenia, C. (2021). *Seputar Fungsi Microsoft Excel yang Perlu Kalian Ketahui*. Dqlab.Id. <https://dqlab.id/seputar-fungsi-microsoft-excel-yang-perlu-kalian-ketahui>
- Gokiladevi, M., Santhoshkumar, S., & Varadarajan, V. (2022). *Machine Learning Algorithm Selection For Chronic Kidney Disease*. 1, 102–115.
- Harahap, Malik, & Imam. (2021). *Klasifikasi Diagnosa Penyakit Jantung menggunakan Algoritma Random Forest*. 3.
- Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 211–220. <https://doi.org/10.33480/jitk.v5i2.1181>
- Hermawati, F. A. (2013). *Data Mining* (P. Christian (ed.)). Andi.
- Hikmah, I. R., & Yasa, R. N. (2021). Perbandingan Hasil Prediksi Diagnosis pada

- Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange. *Jurnal Telematika*, 16(2), 69–76.
- Iriadi, N. (2013). Komparasi Algoritma Klasifikasi Data Mining Dalam Penentuan Resiko Kredit Pada Koperasi Serba Usaha. *Paradigma*, XV(2), 192–204.
- Islam, M., Wu, C., & Nasrin, T. (2018). *Applications of Machine Learning in Fatty Live Disease Prediction*. 166–170. <https://doi.org/10.3233/978-1-61499-852-5-166>
- Khasanah, L. U. (2020). *Exploratory Data Analysis : Pahami Lebih Dalam untuk Siap Hadapi Industri Data*. Dqlab.Id. <https://dqlab.id/data-analisis-machine-learning-untuk-proses-pengolahan-data>
- Khasanah, L. U. (2021). *Kenali 4 Library Python yang Populer untuk Data Science*. Dqlab.Id. <https://dqlab.id/kenali-4-library-python-yang-populer-untuk-data-science>
- Kresna. (2019). *Kelebihan Analisis Regresi Logistik (skripsi dan tesis)*. Konsultasiskripsi.Com. <https://konsultasiskripsi.com/2019/09/30/kelebihan-analisis-regresi-logistik-skripsi-dan-tesis/>
- Kurnianto, E. A., Cholissodin, I., & Santoso, E. (2018). Klasifikasi Penderita Penyakit Ginjal Kronis Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(12), 6597–6602.
- Kuswandi, W. Y., Ichsan, N., Ernawati, E., & Wahyuni, T. (2021). Sistem Informasi Pelayanan Karier Siswa Dan Alumni Berbasis Framework Codeigniter. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 13(2), 12–20. <https://doi.org/10.35969/interkom.v13i2.45>
- Makarim, dr. F. R. (2022). *Fungsi Hati*. Halodoc.Com. <https://www.halodoc.com/kesehatan/fungsi-hati>
- Patimah, E., Haekal, B. V., Prasvita, D. S., Komp, S., & Kom, M. (2021). *Klasifikasi Penyakit Liver dengan Menggunakan Metode Decision Tree*. *April*, 655–659. <https://conference.upnvj.ac.id/index.php/senamika/article/view/1388>

- Pittara, D. (2022). *Penyakit Liver*. Alodokter.Com. <https://www.alodokter.com/penyakit-liver>
- Prabiantissa. (2021). *Klasifikasi pada Dataset Penyakit Hati Menggunakan Algoritma Support Vector Machine, K-NN, dan Naïve Bayes*. 219–224.
- Priya, M. B., Juliet, P. L., & Tamilselvi, P. R. (2018). *Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms*. 206–211.
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *J-Ptiik*, 2(2), 802–810. <http://j-ptiik.ub.ac.id>
- Pusporani, E., Qomariyah, S., & Irhamah, I. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, 2(1), 25. <https://doi.org/10.12962/j27213862.v2i1.6810>
- Restiani, D.-. (2018). Kombinasi Algoritma Cart Dan Ripper Untuk Mendiagnosis Penyakit Liver Berbasis Correlation Based Feature Selection. *Jurnal Teknik Informatika*, 11(1), 31–36. <https://doi.org/10.15408/jti.v11i1.6660>
- Retnoningsih, P. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- Reyvan. (2021). *Belajar Python dengan Google Colab*. Www.Dqlab.Id. <https://www.dqlab.id/belajar-python-dengan-google-colab>
- Rianto, W. (2015). Resampling Logistic Regression untuk Penanganan Ketidakseimbangan Class pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(1), 46–53.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sandag, G. A. (2020). Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITo Smart Journal*, 6(2), 167–178. <https://doi.org/10.31154/cogito.v6i2.270.167-178>
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu.

- Setiawati, I., Wibowo, A. P., & Hermawan, A. (2019). Pendahuluan Tinjauan Pustaka Penelitian Sebelumnya Klasifikasi. *Jurnal of Information System Management*, 1(1), 13–17.
- Singaravelu, M., Krishnan, S., & Karthik, K. (2018). *Classification of Liver Patient Dataset Using Machine Learning Algorithms Classification of Liver Patient Dataset Using Machine Learning Algorithms*. September. <https://doi.org/10.14419/ijet.v7i3.34.19217>
- Soraya, S. I. (2019). *Perbedaan Boxplot, Scatter Plot dan Violin Plot*. Medium.Com. <https://medium.com/@shelvyyy/perbedaan-boxplot-scatter-plot-dan-violin-plot-49eef92a2bd2>
- Suardika, I. G. I. (2019). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi Dan Bisnis Universitas Pendidikan Nasional. *Jurnal Ilmu Komputer Indonesia*, 4(2), 37–44. <https://doi.org/10.23887/jik.v4i2.2775>
- Suhendra, M. A., Ispriyanti, D., & Sudarno, S. (2020). Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera Di Kota Semarang Menggunakan Metode Regresi Logistik Biner Dan Metode Chaid. *Jurnal Gaussian*, 9(1), 64–74. <https://doi.org/10.14710/j.gauss.v9i1.27524>
- Susilowati, E., Sabariah, M. K., Gozali, A. A., Informatika, J. T., Telkom, U., & Machine, S. V. (2015). *Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter Implementation Support Vector Machine Method For Traffic Jam Classification On Twitter*. 2(1), 1478–1484.
- Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D'CARTESIAN*, 6(2), 56. <https://doi.org/10.35799/dc.6.2.2017.17023>
- Trivusi. (2022a). *Algoritma Random Forest: Pengertian dan Kegunaannya*. Trivusi.Web.Id. <https://www.trivusi.web.id/2022/08/algoritma-random-forest.html>
- Trivusi. (2022b). *Penjelasan Lengkap Algoritma Support Vector Machine (SVM)*. Trivusi.Web.Id. <https://www.trivusi.web.id/2022/04/algoritma-svm.html>

- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 134.
- Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 84–92.
- Wijaya, M. (2018). Peningkatan Akurasi pada Algoritma Support Vector Machine dengan Penerapan Information Gain untuk Mendiagnosa Chronic Kidney Disease. *Journal of Physics: Conference Series*, 983(1), 22–27.
- Yogi, I. G., Putra, P., Khrisne, D. C., & Suyadnya, I. M. A. (2019). *Expert System for Early Diagnosis of Heart Disease Using Random Forest Method*. 3(1), 15–18.