

Klasifikasi Penyakit Hepatitis C Menggunakan Algoritma Support Vector Machine

Ni Kadek Trisnawati^{a1}, I Gede Santi Astawa^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹kadektrisna646@gmail.com
²santiastawa@gmail.com

Abstract

Hepatitis is an inflammation of the liver which is often caused by a virus. Hepatitis that lasts less than 6 months is called "acute hepatitis", hepatitis that lasts more than 6 months is called "chronic hepatitis" Hepatitis consists of various types starting from hepatitis A, B, C, D and E. This study will discuss the classification Hepatitis C is classified as blood based (blood donor), suspected blood donor, fibrosis (moderate to severe hepatitis) and cirrhosis (chronic hepatitis). Classification uses the Support Vector Machines (SVM) algorithm with Confusional Matrix testing. The dataset used was obtained from the kaggle.com site with a total of 590 data consisting of 14 features or attributes. This study produces an accuracy of 93%

Keywords: Hepatitis, Support Vector Machines (SVM)

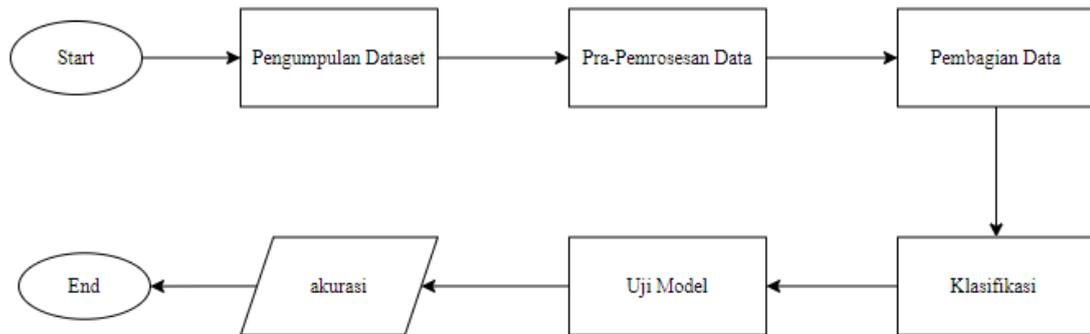
1. Pendahuluan

Penyakit Hepatitis adalah penyakit yang disebabkan oleh beberapa jenis virus yang menyerang dan menyebabkan peradangan serta merusak sel-sel organ hati manusia. Namun, hepatitis juga dapat disebabkan oleh faktor lain seperti autoimun, alkohol, dan obat-obatan (Martinez, 1996). Virus hepatitis C adalah virus yang ditularkan melalui darah dan sebagian besar infeksi terjadi melalui paparan darah dari praktik injeksi yang tidak aman, perawatan kesehatan yang tidak aman, transfusi darah yang tidak diskriminasi, penggunaan narkoba suntikan, dan praktik seksual yang menyebabkan paparan darah [1]. Tingkatan dari hepatitis C ini seperti fibrosis yaitu penyakit hepatitis dengan level sedang berat dan sirosis yaitu penyakit hepatitis dengan level kronis. Dalam masalah ini akan dilakukan klasifikasi jenis penyakit hepatitis C menggunakan algoritma Support Vector Machine (SVM). Algoritma SVM merupakan algoritma yang sudah terbukti efektif dalam klasifikasi data yang kompleks. Dalam kasus klasifikasi sudah terdapat banyak algoritma yang digunakan seperti, Naïve Bayes, Random Forest, K-Nearest Neighbour (KNN) dan Support Vector Machines (SVM). KNN menjadi algoritma yang paling sering digunakan oleh para peneliti dalam bidang klasifikasi. Namun, dalam beberapa penelitian menunjukkan tingkat akurasi SVM lebih tinggi dibandingkan dengan algoritma KNN. Karamoy, Gandis Helen (2021) melakukan tingkat kerusakan hati yang diakibatkan oleh virus hepatitis C. Pada penelitian tersebut menggunakan beberapa metode seperti Support Vector Machine (SVM), Random Forest (RF), Naive Bayes, Logistic Regression dan kNN. Dengan menggunakan dataset dari the University of California Irvine Machine Learning Repository (UCI-MLR), data di dibuat oleh penelitian memaparkan akurasi terbaik dari algoritma Support Vector Machines (SVM) sebesar 0,883 dan akurasi terendah dengan menggunakan algoritma decision tree. Indri Monika Parapat, Muhammad Tanzil Furqon, Sutrisno (2018) melakukan penelitian terhadap penyimpangan tumbuh kembang anak. Pada penelitiannya menggunakan metode SVM dengan 90 data yang dibagi menjadi 3 kelas. Kelas penelitian ini mewakili 3 jenis penyimpangan tumbuh kembang anak yaitu Down Syndrome, Autisme, dan attention deficit hyperactivity disorder (ADHD). Hasil akhir dari penelitian ini menghasilkan rata-rata akurasi tertinggi sebesar 63,11% $\lambda = 10$, $C = 1$, $\text{itermax} = 200$ dan juga menggunakan kernel polynomial. Penelitian ini dilakukan untuk mengklasifikasi penyakit hepatitis C apakah tergolong ke dalam diagnosis blood donor (memiliki riwayat pen donor darah), suspect blood donor (diduga memiliki potensi menjadi pendonor

darah), hepatitis, fibrosis atau sirosis. Dataset yang digunakan berasal dari dataset kaggle “Hepatitis C virus – Blood Based Detection” dengan 14 atribut di dalamnya.

2. Metode Penelitian

Dalam penelitian ini dilakukan dengan menggunakan algoritma SVM, adapun proses pada sistem ini berjalan seperti berikut



Gambar 1. Flowchart Klasifikasi penyakit hepatitis

2.1 Pengumpulan Data

Penelitian ini menggunakan dataset yang didapatkan dari situs kaggle yaitu “Hepatitis C virus – Blood bases Detection” dan disimpan ke dalam google drive. Dataset ini terdiri dari 590 data csv dengan 14 atribut diantaranya yaitu ID, Category(diagnosis), Age (umur), Sex (jenis kelamin), ALB (serum albumin), ALP (fosfatase alkali), ALT (alanine aminotransferase), AST (aspartate aminotransferase), BIL (bilirubin), CHE (kolinesterase), CHOL (kolesterol), CREA kreatini), GGT (gamma-glutamyl transferase), PROT (protein).

2.2 Pra-Pemrosesan Data

Pada tahapan ini dilakukan seleksi fitur, dimana fitur yang tidak digunakan akan dihilangkan atau dihapus dari dataset. Adapun fitur atau atribut yang digunakan dalam penelitian ini yaitu usia (Age), tingkat albumin (ALB), tingkat fosfatase alkali (ALP), tingkat aspartate aminotransferase (AST), dan tingkat bilirubin (BIL). Setelah atribut atau fitur ditentukan selanjutnya melakukan pemisahan data menjadi data training dan data testing. Pemisahan ini menggunakan ‘train_test_split()’ dari scikit-learn yaitu data training (x_train, y_train) dan data testing (x_test, y_test).

2.3 Pembagian Data

Table 1. Pembagian Data

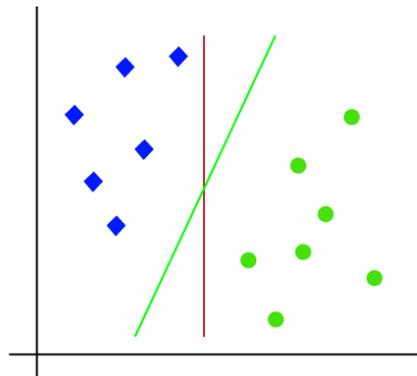
Jumlah Data	Persentase
462	80%
128	20%

2.4 Klasifikasi

a. Support Vector Machines (SVM)

Support Vector Machine atau SVM adalah algoritme pembelajaran mesin yang diawasi yang dapat digunakan untuk klasifikasi dan regresi. Support Vector Machines (SVM)

merupakan algoritma membentuk ruang hitosis berupa fungsi-fungsi linier di sebuah ruang fitur berdimensi tinggi [2]. SVM dibagi menjadi dua jenis yaitu SVM linear dan SVM non-linear, pada penelitian ini menggunakan algoritma SVM linear. SVM linier digunakan pada data yang memiliki kemampuan untuk dipisahkan secara linear. Artinya, jika sebuah dataset dapat dibagi menjadi dua kelas menggunakan sebuah garis lurus tunggal, maka data tersebut dianggap dapat dipisahkan secara linier, dan pengklasifikasi yang digunakan disebut sebagai pengklasifikasi SVM linear. Hyperplane dengan margin maksimum dan margin untuk SVM yang dilatih dengan sampel dari dua kelas. Sampel pada margin disebut vektor pendukung.



Gambar 2. Margin SVM Linear

Adapun rumus Support Vector Machines Linear yang digunakan seperti berikut :

$$y(x) = w^T * x + b$$

Di mana:

- $y(x)$ adalah prediksi kelas untuk sampel x .
- w adalah vektor bobot yang ditemukan selama proses pelatihan SVM linier.
- x adalah vektor fitur dari sampel yang ingin diprediksi.
- b adalah bias (intercept) yang ditentukan selama proses pelatihan SVM linier.

b. Uji Model

Pengujian dilakukan dengan menggunakan confusion matrixs dimana confusion matrix sendiri merupakan alat penting untuk mengevaluasi sejauh mana model klasifikasi mampu melakukan prediksi yang benar. Dengan menganalisis kombinasi dari keempat istilah tersebut, kita dapat memperoleh wawasan tentang seberapa baik model dalam mengklasifikasikan data ke dalam kelas yang benar.

3. Hasil dan Pembahasan

3.1 Confusion Matrix

Pada penelitian ini digunakan untuk memberikan indikasi sejauh mana model klasifikasi SVM dapat memprediksi dengan benar setiap kelas atau kelas yang ada. Dengan menganalisis matriks ini, kita dapat mendeteksi kesalahan klasifikasi yang dibuat oleh model dan mengukur keefektifan dan akurasi model dalam membuat prediksi. Berikut merupakan confusion matrix yang dihasilkan dalam penelitian ini

menerapkan algoritma ini yaitu dengan perhitungan SVM linear menggunakan atribut dari dataset. Pengujian dengan menggunakan confusion matrix menghasilkan akurasi sebesar 93%.

Daftar Pustaka

- [1] World Health Organization. 24 June 2022. Hepatitis C. Diakses pada 09 Juni 2023, dari <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>
- [2] Zuama, RA. 2021. Pembelajaran Mesin untuk diagnosis tingkat kerusakan hati akibat hepatitis C. Sentra Penelitian Engineering dan Edukasi. Volume 13 No 3.
- [3] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- [4] Parapat, IM dkk. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol. 2, No.10
- [5] Nurlaily dkk. (2022). Classification of Hepatitis Patients Using Logistic Regression and Support Vector Machines Methods. *Jurnal Pendidikan Matematika*. Volume 5, Number 2

Halaman ini sengaja dibiarkan kosong