

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Prevalensi gagal jantung global pada 2021 sekitar 55,5 juta kasus dengan 676,7 per 100.000 orang. Di AS, 6,7 juta orang dewasa menderita gagal jantung tahun 2021, diperkirakan naik ke 8,7 juta tahun 2030 dan 11,4 juta tahun 2050. Penyebab utama adalah *penyakit jantung iskemik* (34,53%), *hipertensi* (22,53%), dan *kardiomiopati* (7,61%). Kasus gagal jantung meningkat terutama pada usia muda, minoritas ras dan etnis, serta pasien dengan penyakit penyerta. Mortalitas akibat penyakit kardiovaskular sangat tinggi di AS, dengan satu kematian setiap 34 detik. Gagal jantung fokus utama kesehatan global karena tingginya kesakitan dan kematian, sehingga deteksi dini dan prediksi risiko dengan teknologi seperti *machine learning* sangat diperlukan untuk mengurangi beban sistem kesehatan [1] [2] [3] [4].

*Logistic Regression* dan *Support Vector Machine (SVM)* adalah dua algoritma *machine learning* yang populer dan sering digunakan dalam klasifikasi data klinis. Berdasarkan tiga sumber yang dianalisis, *Logistic Regression* dipilih karena modelnya yang mudah diinterpretasi dan dilaporkan memiliki akurasi kompetitif, dengan satu studi menunjukkan akurasi sekitar 81% dalam prediksi penyakit jantung, meskipun sedikit lebih rendah dibandingkan algoritma lain yang lebih kompleks. Di sisi lain, *SVM* unggul dalam menghadapi data *non-linear* dan berdimensi tinggi dengan kemampuannya menemukan *hyperplane* pemisah yang optimal; akurasinya bervariasi antara 62,64% hingga 90% pada berbagai tugas klasifikasi seperti diagnosis *Alzheimer* dan prediksi penyakit jantung. Dengan demikian, kedua algoritma ini menunjukkan keunggulan masing-masing dengan akurasi yang bersaing, tergantung pada karakteristik data klinis serta metode analisis yang digunakan, sehingga keduanya sangat relevan dalam konteks aplikasi medis modern yang menuntut prediksi dan klasifikasi yang akurat dan andal [5] [6] [7].

Berbagai studi komparatif telah menunjukkan hasil yang bersifat kontekstual dalam membandingkan kedua algoritma ini. *Logistic Regression* biasanya lebih menguntungkan dalam hal interpretabilitas dan penanganan data dengan fitur yang lebih sedikit atau hubungan linier, sementara *SVM* lebih mampu menangani data dengan kompleksitas dan variabilitas tinggi yang umum ditemui dalam dataset klinis

real-world [8] [9] [10]. Tantangan yang sering ditemui dalam aplikasi kedua algoritma ini meliputi pemilihan fitur yang tepat, teknik *preprocessing* data, serta optimasi parameter yang berpengaruh besar terhadap hasil klasifikasi. Hal ini menggarisbawahi kebutuhan untuk evaluasi berkelanjutan yang berbasis data klinis mutakhir agar dapat memastikan prediksi yang tepat dan relevan dalam konteks kesehatan pasien [11] [12] [13].

Lebih jauh lagi, perbandingan kinerja Logistic Regression dan SVM dalam domain prediksi gagal jantung tidak hanya memberikan gambaran akurasi dari sisi teknis, tetapi juga berkontribusi pada pengembangan sistem pendukung keputusan medis yang efektif. Model yang dihasilkan diharapkan dapat mendukung praktisi medis dalam melakukan diagnosis lebih cepat dan membuat keputusan terapeutik yang lebih tepat, sehingga meningkatkan efektivitas pengobatan dan menurunkan angka rawat inap pasien gagal jantung [14]. Penelitian ini juga membuka kesempatan untuk mengintegrasikan algoritma lain serta menerapkan teknik *ensemble* yang berpotensi menghasilkan performa lebih baik di masa depan.

Dengan cara ini, dapat diperoleh pengetahuan yang lebih komprehensif tentang kekuatan dan kelemahan dari setiap algoritma dalam konteks prediksi gagal jantung, yang tidak hanya bermanfaat secara akademis, tetapi juga praktis dalam implementasi di dunia kesehatan [15] [13].

## **1.2. Rumusan Masalah**

Permasalahan yang dikaji dalam penelitian ini adalah bagaimana membandingkan efektivitas algoritma *Logistic Regression* dan *Support Vector Machine (SVM)* dalam memprediksi gagal jantung menggunakan dataset klinis, dengan evaluasi kinerja berdasarkan metrik akurasi, presisi, recall, dan *F1-score*.

## **1.3. Tujuan dan Manfaat**

### **1.3.1. Tujuan**

Penelitian ini memiliki tujuan untuk melakukan perbandingan kinerja antara algoritma *Logistic Regression* dan *Support Vector Machine (SVM)* dalam memprediksi gagal jantung dengan memanfaatkan dataset klinis yang ada. Analisis dilakukan dengan mengukur berbagai metrik evaluasi seperti akurasi, presisi, recall, dan *F1-score* agar dapat menentukan algoritma yang paling efektif dan efisien untuk aplikasi prediksi di

bidang klinis. Dengan demikian, penelitian ini memberikan fokus yang jelas pada studi perbandingan performa dua algoritma pembelajaran mesin yang banyak digunakan dalam kesehatan, sekaligus menyediakan dasar untuk pemilihan algoritma terbaik dalam konteks tersebut.

### **1.3.2. Manfaat**

Harapan dari penelitian ini mencakup manfaat-manfaat berikut:

1. Menyediakan perbandingan kinerja yang jelas antara algoritma *Logistic Regression* dan *Support Vector Machine (SVM)* dalam memprediksi gagal jantung menggunakan data klinis.
2. Mendukung pengembangan sistem pendukung keputusan medis yang lebih akurat dan andal, sehingga membantu dokter dalam diagnosis dini dan pengambilan keputusan terapeutik yang tepat, yang pada akhirnya dapat meningkatkan kualitas perawatan pasien gagal jantung.
3. Mengidentifikasi algoritma yang paling efisien dan efektif, sehingga memberikan dasar yang kuat untuk implementasi praktis dalam aplikasi klinis serta mendorong inovasi di bidang teknologi kesehatan.

### **1.4. Batasan Masalah**

Penelitian ini akan terfokus dalam mengklasifikasikan tingkat kecemasan dengan batasan-batasan masalah sebagai berikut:

1. Studi ini fokus pada analisis menggunakan dataset klinis yang ada dan sesuai untuk prediksi gagal jantung, sehingga cakupan hasil penelitian terbatas oleh karakteristik data yang digunakan.
2. Analisis performa kedua algoritma difokuskan pada metrik evaluasi akurasi, presisi, recall, dan *F1-score* tanpa mempertimbangkan metrik lain seperti *ROC-AUC* atau waktu komputasi.
3. Penelitian ini membatasi penggunaan metode pembelajaran mesin pada dua algoritma yaitu *Logistic Regression* dan *Support Vector Machine (SVM)*, tanpa memasukkan algoritma lain atau teknik *ensemble* yang mungkin juga efektif.
4. Hasil penelitian hanya berlaku pada konteks prediksi gagal jantung dan tidak serta-merta dapat digeneralisasi ke prediksi penyakit lain atau aplikasi medis lainnya.
5. Dataset dipakai berbentuk csv yang diperoleh dari

## 1.5. Keterbaruan

Tabel 1.1 Keterbaruan

No	Penelitian Sebelumnya	Keterbaruan
1.	Prabowo et al., (2023), Penerapan dan perbandingan kinerja empat algoritma <i>machine learning</i> , dalam mendeteksi penyakit jantung menggunakan dataset dengan 14 fitur klinis yang diambil dari Kaggle. Penelitian ini menunjukkan bahwa ketiga algoritma <i>SVM</i> , <i>RF</i> , dan <i>AdaBoost</i> mencapai kinerja sangat tinggi dengan akurasi sebesar 98,54%, sedangkan <i>Logistic Regression</i> memiliki akurasi lebih rendah, yaitu sekitar 80,98%.	Keterbaruan utama terletak pada analisis komparatif yang mendalam dan konsisten menggunakan metrik lengkap (akurasi, <i>precision</i> , recall, <i>F1-score</i> ) yang menegaskan superioritas algoritma <i>ensemble</i> dan <i>SVM</i> dibandingkan <i>LR</i> dalam konteks prediksi penyakit jantung, sekaligus menggarisbawahi pentingnya pemilihan algoritma yang sesuai berdasarkan karakteristik dataset [15].
2.	Fabiyanto et al., (2024), Penelitian ini menyajikan dan membandingkan performa dua model pembelajaran mesin, yakni <i>Logistic Regression</i> ( <i>LR</i> ) dan <i>Support Vector Machine</i> ( <i>SVM</i> ), dalam memprediksi penyakit jantung dengan menggunakan dataset UCI Heart Disease yang berisi 14 fitur klinis. Hasil penelitian menunjukkan bahwa <i>Logistic Regression</i> memperoleh akurasi tertinggi sebesar 89%, serta mencatat performa terbaik pada metrik presisi, recall, dan <i>F1-score</i> , sementara <i>SVM</i> mencatat akurasi mencapai 85%.	Keterbaruan penelitian ini terletak pada penerapan evaluasi komprehensif kedua algoritma menggunakan berbagai metrik performa yang secara jelas menegaskan keunggulan <i>Logistic Regression</i> dalam klasifikasi penyakit jantung, sekaligus memberikan dasar yang kuat untuk pemilihan model yang efektif di lingkungan klinis [16].
3.	Rimal et al., (2025), Penggunaan perbandingan komprehensif antara algoritma <i>Logistic Regression</i> ( <i>LR</i> ) dan <i>Support Vector Machine</i> ( <i>SVM</i> ) dengan kernel yang dioptimalkan untuk prediksi penyakit jantung menggunakan dataset dengan 13 fitur klinis yang telah dinormalisasi.	Hasil penelitian menunjukkan bahwa <i>Logistic Regression</i> sedikit lebih unggul dengan akurasi mencapai 83,81%, sedangkan <i>Support Vector Machine</i> ( <i>SVM</i> ) memperoleh akurasi sebesar 82,49%. Temuan ini menegaskan pentingnya pemilihan algoritma yang tepat dengan mempertimbangkan keseimbangan antara akurasi dan efisiensi model dalam aplikasi prediksi klinis penyakit jantung. [4].