BAB II
TINJAUAN PUSTAKA

2.1 Penelitian yang Relevan

Penelitian yang relevan ini dilakukan untuk membandingkan penelitian sebelumnya dari jurnal dan paper yang relevan sebagai referensi topik penelitian untuk memperkuat posisi penelitian ini sendiri. Ini adalah beberapa contoh penelitian sebelumnya.

Penelitian pertama yang diteliti oleh Yohannes, Muhammad Rizky Pribadi, dan Leo Chandra pada tahun 2020 berjudul "*Klasifikasi Jenis Buah dan Sayuran Menggunakan SVM dengan Fitur Saliency-HOG dan Color Moments*" membicarakan cara menggabungkan fitur bentuk dan warna untuk mengklasifikasikan gambar buah dan sayuran. Fitur bentuk diperoleh melalui Saliency-HOG, sementara informasi warna diperoleh dengan menggunakan Color Moments. Dataset yang digunakan adalah Fruit 360, yang terdiri dari 11.400 citra dari 114 jenis buah dan sayuran, dibagi dengan proporsi 70% untuk data latih dan 30% untuk data uji. Peneliti menguji empat metode saliency (FT, HC, RC, SR) untuk melihat perbedaan hasil ekstraksi. Model klasifikasi dibangun menggunakan algoritma SVM, dan evaluasi dilakukan berdasarkan precision, recall, dan akurasi. Hasilnya menunjukkan bahwa kombinasi FT-HOG dan Color Moments memberikan hasil terbaik, dengan precision 98,57%, recall 98,55%, dan akurasi 99,12%. Hasil dari penelitian ini menunjukkan bahwa gabungan fitur saliency dan warna sangat efektif untuk klasifikasi buah dan sayuran, dengan SVM menghasilkan akurasi yang tinggi (Yohannes et al., 2020).

Penelitian kedua yang diteliti oleh Laili Nur Farida dan Saiful Bahri pada tahun 2024 berjudul "*Klasifikasi Gagal Jantung Menggunakan Metode SVM (Support Vector Machine)*" membahas penerapan algoritma SVM yang digunakan untuk mengklasifikasikan gagal jantung berdasarkan data klinis pasien. Dataset yang digunakan berasal dari Kaggle, terdiri dari 918 data pasien dengan 11 atribut prediktor seperti usia, jenis kelamin, tekanan darah, kolesterol, detak jantung maksimum, dan atribut medis lainnya, serta satu atribut target yaitu status gagal jantung. Data diproses melalui tahapan preprocessing, meliputi konversi data kategorikal ke numerik, deteksi

outlier menggunakan metode IQR, dan normalisasi z-score. Setelah itu, data dibagi menjadi data latih dan data uji dengan menggunakan perbandingan 90:10. Model klasifikasi dibuat dengan menggunakan solver variabel diskret (SVM) dengan linear, RBF, dan polinomial, serta dilakukan pengujian dengan variasi parameter cost. Evaluasi kinerja model dilakukan dengan memanfaatkan nilai F1, akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa kernel linear memiliki kinerja terbaik, dengan akurasi tertinggi sebesar 86,92%. Model terbaik menghasilkan nilai presisi 88,68%, recall 87,85%, dan F1-score 88,26%, yang membuktikan bahwa metode SVM sangat baik dalam mengklasifikasikan penyakit gagal jantung dan berpotensi membantu deteksi dini secara lebih akurat (Farida & Bahri, 2024).

Penelitian ketiga yang diteliti oleh Elizabeth Fiona Hartono dan Nur Rachmat (2022) dalam penelitian berjudul "*Klasifikasi Jenis Plastik HDPE, LDPE, dan PS Berdasarkan Tekstur Menggunakan Metode Support Vector Machine*" membahas klasifikasi jenis plastik berdasarkan tekstur citra. Penelitian ini menggunakan ekstraksi ciri *Local Binary Pattern* (LBP) dan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi. Dataset terdiri dari 180 citra plastik berjenis HDPE, LDPE, dan PS dengan pembagian 70% data latih dan 30% data uji, serta ukuran citra 450 × 450 piksel. Proses preprocessing meliputi pemotongan citra dan konversi RGB ke grayscale. Klasifikasi dilakukan menggunakan kernel RBF, linear, dan polynomial, dengan evaluasi berupa akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa kernel polynomial memiliki akurasi dan kinerja terbaik 95,05%, presisi 92,78%, dan recall 92,58%, sehingga kombinasi LBP dan SVM efektif untuk klasifikasi jenis plastik berdasarkan tekstur (Hartono & Rachmat, 2022).

Penelitian keempat yang diteliti oleh Yoseph P. K. Kelen dan Budiman Baso pada tahun 2023 berjudul "*Klasifikasi Tenun Timor Menggunakan Metode SVM Berdasarkan Fitur SURF dengan Representasi Bag of Visual Words*" membahas klasifikasi citra motif tenun Timor sebagai upaya pelestarian budaya melalui pengolahan citra digital. Penelitian ini menggunakan ekstraksi fitur *Speeded Up Robust Features* (SURF) yang direpresentasikan dengan *Bag of Visual Words* (BoVW), sedangkan proses klasifikasi dilakukan menggunakan *Support Vector Machine* (SVM) multi-kelas. Dataset yang digunakan terdiri dari 420 citra tenun

Timor dengan 7 kelas motif, yang dipisahkan menjadi data uji dan data latih menggunakan 5-fold cross validation. Peneliti menguji variasi jumlah cluster BoVW (100, 300, dan 500) serta kernel SVM (linear, polynomial, dan RBF) dengan pendekatan OVA dan OVO. Hasil penelitian menunjukkan bahwa SVM kernel linear dengan metode OVA dan jumlah cluster 500 memberikan kinerja optimal dengan akurasi sebesar 98,10%, sehingga kombinasi SURF, BoVW, dan SVM terbukti efektif untuk klasifikasi citra motif tenun Timor (Kelen & Baso, 2023).

Penelitian kelima yang dilakukan oleh Yusuf Amrozi beserta timnya pada tahun 2022, dengan judul "*Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM*", menjelaskan implementasi algoritma Support Vector Machine (SVM) untuk mengidentifikasi varietas pisang melalui analisis warna gambar. Data yang dipakai mencakup 1.256 foto pisang, yang dikategorikan ke dalam dua jenis utama, yaitu Pisang Ambon dan Pisang Lady Finger. Dari uji coba yang dilakukan, metode SVM berhasil mencapai tingkat keakuratan sebesar 89,86%, dengan nilai *True Positive* (TP) mencapai 0,82, *False Positive* (FP) sebesar 0,18, *False Negative* (FN) hanya 0,02, serta *True Negative* (TN) sebesar 0,98. Temuan ini mengindikasikan bahwa SVM terbukti ampuh untuk membedakan jenis pisang berdasarkan nuansa warna pada citra (Amrozi et al., 2022).

Penelitian keenam yang diteliti oleh Mareanus Lase, Windu Gata, Hiya Nalatissifa, dan Sri Diantika pada tahun 2020 berjudul "*Komparasi Algoritma SVM dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik*" membahas perbandingan algoritma SVM dan Naive Bayes, dalam mengklasifikasikan kestabilan jaringan listrik. Dataset yang digunakan berasal dari *Electrical Grid Stability Simulated* dengan 10.000 data dan 14 atribut. Hasilnya menunjukkan bahwa SVM mencapai akurasi 98,9%, sementara *Naive Bayes* hanya 97,64%. Meskipun hasil pengujian dengan berbagai konfigurasi data training dan testing menunjukkan variasi, SVM secara konsisten menunjukkan kinerja lebih baik. Hasil Studi ini menunjukkan SVM lebih baik untuk klasifikasi kestabilan jaringan listrik dan lebih direkomendasikan untuk aplikasi serupa (Diantika et al., 2021).

Penelitian ketujuh yang diteliti oleh Oryza Habibie Rahman dan tim pada tahun 2021 berjudul "*Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine*" membahas penggunaan metode SVM untuk mengklasifikasikan ujaran kebencian di Twitter ke dalam lima kategori: suku, agama, ras, antar golongan, dan netral. Dataset yang digunakan terdiri dari kalimat-kalimat teridentifikasi sebagai ujaran kebencian, yang dibagi menjadi data latih dan uji. Peneliti menguji tiga jenis kernel SVM: linear, sigmoid, dan Radial Basis Function (RBF). Hasilnya menunjukkan bahwa kernel RBF memberikan akurasi terbaik, yaitu 93%, lebih tinggi dibandingkan dengan kernel linear dan sigmoid. Hasil penelitian ini menyimpulkan bahwa SVM dengan kernel RBF mengklasifikasikan dengan baik ujaran kebencian di Twitter dan dapat digunakan untuk menganalisis perilaku masyarakat online terkait isu-isu sensitive (Rahman et al., 2021).

Penelitian kedelapan yang dilakukan oleh Taufik Hidayatulloh dan Lestari Yusuf pada tahun 2023, dengan judul "*Klasifikasi Tipe Berat Tubuh Menggunakan Metode Support Vector Machine*", menguraikan penggunaan algoritma Support Vector Machine (SVM) dalam mengelompokkan jenis berat badan menjadi dua kategori, yakni normal dan overweight. Studi ini memanfaatkan 252 sampel data pria yang diambil dari dataset publik Kaggle, yang mencakup fitur antropometri dan diberi label sesuai dengan indeks massa tubuh (BMI), dengan pembagian data sebesar 70% untuk pelatihan dan 30% untuk pengujian. Dari analisis yang dilakukan, metode SVM menunjukkan tingkat akurasi sebesar 92,11% serta nilai AUC mencapai 0,990, sehingga dapat ditarik kesimpulan bahwa algoritma SVM terbukti sangat berguna untuk mengklasifikasikan tipe berat tubuh (Hidayatulloh & Yusuf, 2023).

Penelitian kesembilan yang diteliti oleh Al Danny R. Wibisono, Eka P. Mandyartha, dan Muhammad M. Al Haromainy pada tahun 2025 berjudul "*Klasifikasi Penyakit Kulit Berbasis Support Vector Machine Dengan Ekstraksi Fitur ABCD Rule*" membahas penggunaan algoritma SVM untuk mengklasifikasikan penyakit kulit dengan menganalisis fitur ABCD Rule. Dataset yang digunakan berisi gambar penyakit kulit dari platform Kaggle, dengan proses meliputi pengumpulan data, pre-processing, segmentasi gambar, dan ekstraksi fitur seperti asimetri, border, warna, dan diameter. Hasil eksperimen menunjukkan akurasi 86,42%, sensitivitas

86,42%, dan spesifisitas 96,60%, dengan penggunaan kernel *Radial Basis Function* (RBF) dan pengaturan parameter yang optimal. Hasil penelitian ini menunjukkan bahwa SVM dengan analisis fitur ABCD Rule efektif dalam mengklasifikasikan penyakit kulit, serta berpotensi meningkatkan efisiensi dan keterjangkauan diagnosis (Wibisono et al., 2025).

Penelitian kesepuluh yang diteliti oleh Eka Sri Rahayu dan tim pada 2025 berjudul "*Klasifikasi Penyakit Jamur Pada Tanaman Tomat dengan Algoritma SVM*" membicarakan penggunaan algoritma *Support Vector Machine* (SVM) untuk mengidentifikasi penyakit pada tanaman tomat. Dataset yang digunakan mencakup 600 gambar daun tomat, yang dikelompokkan ke dalam tiga kategori: sehat, virus mosaik, dan virus ikal daun kuning. Setelah melalui tahapan preprocessing, seperti normalisasi gambar dan pembagian dataset, model SVM kemudian dilatih dan diuji. Hasil penelitian menunjukkan ketepatan sebesar 98,33%, yang menegaskan bahwa SVM berfungsi dalam mendeteksi penyakit pada tanaman tomat (Sri Rahayu et al., 2025).

Tabel 2. 1 Penelitian Yang Relevan

No	Penulis	Yohannes, Muhammad Rizky Pribadi, dan Leo Chandra
1	Keterkaitan Penelitian	Sama-sama menggunakan SVM untuk mengklasifikasikan.
	Gap Penelitian	Konteks Objek: Penelitian ini fokus pada klasifikasi citra buah dan sayuran, sementara penelitian saya berfokus pada klasifikasi status kesehatan ginjal. Variabel: Penelitian ini menggunakan fitur bentuk dan warna citra, sementara penelitian saya menggunakan data medis.
	Hasil Penelitian	Penelitian ini menunjukkan bahwa menggunakan SVM memaksimalkan hasil dengan akurasi 99,12%, precision 98,57%, dan recall 98,55% dalam klasifikasi buah dan sayuran dengan gabungan fitur saliency dan warna.
	Tahun	2020
2	Penulis	Laili Nur Farida dan Saiful Bahri

	Keterkaitan Penelitian	Sama-sama menggunakan SVM untuk mengklasifikasikan dan juga sama-sama menggunakan data medis.
	Gap Penelitian	Gap penelitian ini terletak pada fokus yang berbeda, penelitian ini fokus pada klasifikasi penyakit gagal jantung, sedangkan penelitian saya fokus pada status kesehatan ginjal.
	Hasil Penelitian	Hasilnya menunjukkan bahwa SVM kernel linear memberikan performa dan akurasi terbaik sebesar 86,92%.
	Tahun	2024
3	Penulis	Elizabeth Fiona Hartono dan Nur Rachmat
	Keterkaitan Penelitian	Keterkaitan antara kedua penelitian terletak pada penggunaan model *support vector machine* (svm) untuk analisis data.
	Gap Penelitian	Gap penelitian terletak pada fokus yang berbeda, penelitian ini berfokus pada klasifikasi jenis plastik, sementara penelitian saya berfokus pada klasifikasi status kesehatan ginjal.
	Hasil Penelitian	SVM dengan kernel polynomial memberikan kinerja terbaik dengan akurasi 95,05%, presisi 92,78%, dan recall 92,58%.
	Tahun	2022
4	Penulis	Yoseph P. K. Kelen dan Budiman Baso
	Keterkaitan Penelitian	Keterkaitan antara kedua penelitian adalah penggunaan metode *support vector machine* (svm) untuk melakukan klasifikasi.
	Gap Penelitian	Gap penelitian ini terletak pada konteks berbeda, penelitian ini fokus pada klasifikasi Tenun Timor, sedangkan penelitian saya fokus pada klasifikasi status kesehatan ginjal.
	Hasil Penelitian	penelitian tersebut menunjukkan bahwa svm dengan kernel linear memberikan performa terbaik dengan akurasi sebesar 98,10%.
	Tahun	2023
5	Penulis	Yusuf Amrozi, Dian Yuliati, Agung Susilo, Nur Novianto, dan Rikza Ramadhan

	Keterkaitan Penelitian	Keduanya menggunakan *support vector machine* untuk klasifikasi.
	Gap Penelitian	Penelitian ini berfokus jenis pisang, sementara penelitian saya fokus pada status kesehatan ginjal. Penelitian saya fokus pada masalah medis, sedangkan penelitian sebelumnya fokus pada citra warna.
	Hasil Penelitian	Penelitian ini fokus pada klasifikasi jenis pisang berdasarkan citra warna. Hasil penelitian menunjukkan bahwa svm memiliki akurasi sebesar 89,86% dengan nilai evaluasi lainnya seperti *True Positive* (TP) 0,82, *False Positive* (FP) 0.18, *False Negative* (FN) 0,02, dan *True Negative* (TN) 0,98.
	Tahun	2022
6	Penulis	Sri Diantika, Windu Gata, Hiya Nalatissifa, dan Mareanus
	Keterkaitan Penelitian	Keduanya menggunakan *support vector machine* untuk mengklasifikasi meskipun fokusnya berbeda.
	Gap Penelitian	Gap penelitian terletak pada fokus yang berbeda , penelitian ini berfokus pada klasifikasi kestabilan jaringan listrik, sementara penelitian saya fokus pada klasifikasi status kesehatan ginjal. Selain itu penelitian membandingkan svm dan naïve bayes, sedangkan penelitian saya berfokus pada svm saja.
	Hasil Penelitian	Penelitian tersebut membandingkan 2 algoritma untuk klasifikasi kestabilan jaringan listrik, svm dan *naive bayes*. Hasil penelitian menunjukkan svm mencapai akurasi 98,9% yang berarti lebih efektif dibandingkan *naive bayes* hanya memperoleh 97,64%.
	Tahun	2020
7	Penulis	Oryza Habibie Rahman, Gunawan Abdillah, dan Agus Komarudin
	Keterkaitan Penelitian	Sama-sama menggunakan algoritma svm dalam melakukan klasifikasi.

	Gap Penelitian	Penelitian ini membahas ujaran kebencian pada media sosial twitter, sementara penelitian saya berfokus pada status kesehatan ginjal. Penelitian saya menggunakan variabel klinis, sedangkan penelitian ini menggunakan data kategori (suku, agama, ras, antar golongan, dan netral).
	Hasil Penelitian	Hasil menunjukkan bahwa SVM dengan kernel RBF memberikan akurasi terbaik sebesar 93% lebih tinggi dibandingkan dengan kernel linear dan sigmoid.
	Tahun	2021
8	Penulis	Taufik Hidayatulloh dan Lestari Yusuf
	Keterkaitan Penelitian	keterkaitan penelitian ini terletak pada penggunaan metode yang sama yaitu *support vector machine*.
	Gap Penelitian	penelitian ini fokus pada klasifikasi tipe berat tubuh, sedangkan penelitian saya berfokus pada klasifikasi status kesehatan ginjal.
	Hasil Penelitian	penelitian ini membuktikan bahwa SVM memberikan hasil akurasi sebesar 92,11%.
	Tahun	2023
9	Penulis	Al Danny R. Wibisono, Eka P. Mandyartha, dan Muhammad M. Al Haromainy
	Keterkaitan Penelitian	sama sama klasifikasi berdasarkan data medis menggunakan metode SVM
	Gap Penelitian	Penelitian ini fokus pada klasifikasi penyakit kulit dengan fitur ABCD Rule, Sementara itu, penelitian saya berfokus pada klasifikasi status kesehatan ginjal menggunakan indikator data medis klinis.
	Hasil Penelitian	Hasil ini menunjukkan SVM dengan kernel RBF memiliki akurasi yang baik dalam klasifikasi penyakit kulit dengan fitur ABCD Rule. Akurasi menunjukkan sebesar 86,42% sensitivitas 86,42%, dan spesifisitas 96,60%.

	Tahun	2025
10	Penulis	Eka Sri Rahayu, Oktaviana Anugrah Ade Purnama, Hadi Zakaria, dan Perani Rosyani
	Keterkaitan Penelitian	Keterkaitan antara kedua penelitian adalah menggunakan metode klasifikasi *Support Vector Machine* (SVM).
	Gap Penelitian	Gap penelitian terletak pada jenis data digunakan, dimana penelitian saya menggunakan data medis klinis, sedangkan penelitian tersebut menggunakan data citra gambar.
	Hasil Penelitian	Hasil penelitian tersebut menunjukkan bahwa SVM berhasil mengklasifikasikan penyakit jamur pada tanaman tomat dengan akurasi 98,33% yang menunjukkan akurasi yang baik.
	Tahun	2025

Berdasarkan perbandingan hasil penelitian sebelumnya pada tabel 2.1, dapat disimpulkan bahwa sebagian besar penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) dapat mengklasifikasikan dengan baik berbagai jenis tugas klasifikasi, baik itu menggunakan data citra maupun data numerik. SVM telah digunakan secara efektif dalam banyak bidang, seperti klasifikasi citra buah dan sayuran, jenis plastik, motif tenun, penyakit kulit, penyakit tanaman, ujaran kebencian, kestabilan jaringan listrik, serta untuk menganalisis kondisi medis seperti gagal jantung dan tipe berat tubuh. Hasil-hasil dari penelitian tersebut menunjukkan bahwa pemilihan kernel yang tepat, seperti kernel linear, polinomial, atau *Radial Basis Function* (RBF), serta pengaturan parameter dan teknik preprocessing yang tepat sangat mempengaruhi akurasi, presisi, dan recall model. Selain itu, beberapa penelitian juga menyoroti pentingnya penggunaan *cross-validation* untuk mendapatkan hasil yang lebih akurat dan stabil dalam pengujian model.

Namun, meskipun penelitian-penelitian tersebut sudah memberikan banyak wawasan, ada beberapa aspek yang masih belum sepenuhnya dibahas. Salah satunya adalah penerapan *Support Vector Machine* (SVM) untuk klasifikasi status kesehatan ginjal dengan menggunakan data medis klinis. Sebagian besar penelitian sebelumnya

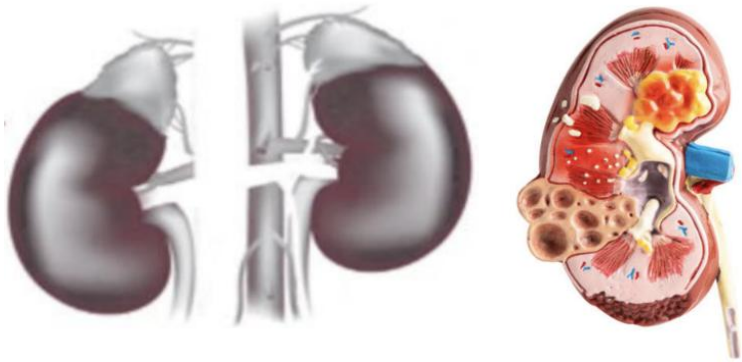
lebih banyak berfokus pada penyakit lain seperti gagal jantung, penyakit kulit, atau diabetes, serta menggunakan dataset citra atau data yang berasal dari luar negeri. Oleh karena itu, penelitian ini memiliki peluang untuk memberikan kontribusi baru dengan mengimplementasikan SVM dalam klasifikasi status kesehatan ginjal berdasarkan data medis klinis, yang diharapkan dapat mendukung deteksi dini dan pengambilan keputusan medis yang lebih tepat.

2.2 Landasan Teori

Berbagai konsep dan teori yang mendasari penelitian ini dijelaskan dalam landasan teori. Teori-teori ini diperoleh dan dikumpulkan dari berbagai sumber yang dapat diandalkan, seperti artikel, buku, dan jurnal.

2.2.1 Ginjal

Organ berbentuk kacang yang disebut ginjal terletak di sisi kanan dan kiri tulang belakang, tepat di bawah hati dan limpa. Ukurannya kurang lebih sebesar kepalan tangan, dengan panjang sekitar 10-12 cm, lebar 5-6 cm, dan ketebalan 3-4 cm, serta berat rata-rata ±150 gram. Sebagai organ pembuangan yang memiliki peran sangat penting, ginjal dilengkapi dengan jutaan unit penyaring berukuran mikroskopis. Unit ini berfungsi menyaring darah untuk mengeluarkan sisa metabolisme, kelebihan mineral, dan cairan yang tidak diperlukan tubuh, kemudian mengeluarkannya dalam bentuk urine (Tapan, 2023).



Gambar 2. 1 Organ Ginjal & Ginjal CKD (Tapan, 2023)

Chronic Kidney Disease (CKD) adalah kondisi di mana terjadi kerusakan pada ginjal, baik dari segi struktur maupun fungsi, yang berlangsung lebih dari tiga bulan.

Penyakit ini menyebabkan penurunan fungsi ginjal secara bertahap, yang bisa disertai dengan penurunan *Laju Filtrasi Glomerulus* (LFG). Meskipun begitu, CKD juga bisa terjadi meski ginjal tidak mengalami kerusakan yang tampak, asalkan laju filtrasi glomerulusnya kurang dari 60 ml/menit/1,73 m2. Penyakit ginjal *Chronic Kidney Disease* (CKD) sering kali tidak menunjukkan gejala di awal, sehingga banyak penderita yang tidak menyadari adanya masalah pada ginjal mereka. Penyakit ini dapat berkembang menjadi gagal ginjal jika tidak ditangani segera, yang memerlukan perawatan medis lebih lanjut, seperti cuci darah atau transplantasi ginjal. Namun, penyakit ginjal bisa dicegah dan dikelola dengan lebih baik jika terdeteksi sejak dini. Deteksi awal memungkinkan pengobatan yang lebih efektif, yang dapat memperlambat atau bahkan menghentikan perkembangan penyakit, serta mengurangi risiko komplikasi berat yang bisa timbul akibat gagal ginjal. Dua penyebab utama penyakit *Chronic Kidney Disease* (CKD) adalah diabetes dan hipertensi, yang seringkali saling berhubungan dan memberikan dampak besar pada kesehatan ginjal. Namun, selain kedua kondisi ini, ada faktor-faktor lain yang juga dapat menyebabkan ginjal *Chronic Kidney Disease* (CKD) (Kyneissia Gliselda, 2021).

2.2.2 Machine Learning

Machine Learning atau komputer dapat belajar sendiri dari data tanpa instruksi langsung. Ini adalah bagian dari teknologi kecerdasan buatan (AI). Dengan menggunakan algoritma tertentu, mesin dapat mengenali pola, tren, atau hubungan dalam data sehingga mampu mengambil keputusan atau memberikan prediksi yang lebih tepat (Asri et al., 2024). Ada berbagai metode dalam *machine learning*, seperti *supervised learning, unsupervised learning, dan reinforcement learning*, yang masing-masing memiliki cara yang berbeda untuk menggunakan data untuk melatih model. Secara keseluruhan, *machine learning* membantu membuat keputusan lebih cepat dan tepat, serta dapat menyesuaikan diri dengan data baru yang terus berkembang (Ranti et al., 2022). Dalam pembelajaran mesin, komputer memanfaatkan algoritma untuk menganalisis data, menemukan pola, dan membuat prediksi. Oleh karena itu, proses ini sangat bergantung pada data yang tersedia. Tanpa adanya data yang cukup dan relevan, algoritma *machine learning* tidak akan dapat

bekerja dengan efektif. Data berperan sebagai elemen utama yang memungkinkan model pembelajaran mesin untuk mengenali pola dan meningkatkan kemampuannya seiring berjalannya waktu (Juliani & Soleh, 2024).

2.2.3 Klasifikasi

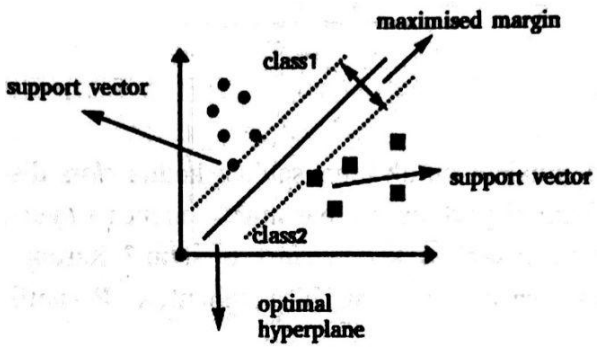
Proses mengelompokkan data atau barang ke dalam kategori yang sudah ditentukan sebelumnya dikenal sebagai klasifikasi, berdasarkan ciri-ciri atau atribut yang dimilikinya. Kategori tersebut biasanya telah ditetapkan sebelumnya sesuai dengan tujuan atau masalah yang ingin diselesaikan. Proses klasifikasi ini sangat berguna untuk membedakan berbagai kelas. Evaluasi dalam klasifikasi sangat penting karena berfungsi untuk menilai sejauh mana kinerja model yang telah dibuat. Melalui evaluasi, kita dapat mengetahui seberapa akurat model dalam memprediksi kategori yang benar (Jatnika et al., 2025).

Klasifikasi merupakan teknik dalam *supervised learning* yang bertujuan untuk meningkatkan akurasi dan keandalan hasil yang diperoleh dari data yang ada. Beberapa model yang sering digunakan dalam klasifikasi antara lain *k-Nearest Neighbor (k-NN), Naive Bayes, Support Vector Machine (SVM), Boosting, Decision Tree, dan Random Forest*. Semua model memiliki kelebihan dan kekurangan, tergantung pada jenis data dan tujuan analisis yang ingin dicapai. Akurasi model klasifikasi dapat dilihat dari berbagai metrik, seperti *accuracy*, *sensitivity* atau *recall*, dan *precision*, yang semuanya dihitung dengan menggunakan *Confusion Matrix* (Nata & Suparmadi, 2022).

2.2.4 Support Vector Machine (SVM)

Salah satu algoritma pembelajaran mesin yang berfokus pada klasifikasi data adalah *Support Vector Machine (SVM)*. Cara kerjanya mirip melalui bagaimana kita memisahkan kelompok data dalam kehidupan sehari-hari, namun dijalankan oleh komputer secara otomatis. SVM berfungsi untuk membantu komputer mengenali pola dalam data dan menempatkan setiap data ke dalam kategori yang tepat. Algoritma ini termasuk dalam *supervised learning*, artinya SVM membutuhkan data yang sudah diberi label sebelumnya untuk belajar. Salah satu keunggulan SVM adalah

kemampuannya menangani data yang tidak bisa dipisahkan secara linear. Untuk itu, SVM menggunakan teknik yang disebut kernel, yang memetakan data ke ruang yang lebih besar sehingga garis pemisah (*hyperplane)* dapat ditemukan. Proses ini memerlukan pengaturan beberapa parameter khusus yang disebut *hyperparameter* (Asri et al., 2025).



Gambar 2. 4 (Primartha & Prof, 2021)

Pada gambar 2.2 menunjukkan s*upport vector, maximised margin*, dan optimal *hyperplane*. Gambar tersebut dapat dilihat bahwa *support vector* ”menyentuh” *boundary* (garis putus-putus) setiap *class*. Di gambar tersebut ada dua buah *support vectors*, kemudikan *hyperplane* berada tepat ditengah dua buah *boundary*.

Beberapa kernel yang sering digunakan adalah linear kernel, polynomial kernel, *radial basis function* (RBF) kernel, dan sigmoid kernel. Kernel ini memungkinkan SVM memetakan data ke ruang yang lebih besar, yang memudahkan pemisahan data yang lebih kompleks, baik yang bersifat linier maupun non-linier. Ini yang menjadikan SVM sebagai metode yang sangat fleksibel dan efektif dalam menyelesaikan masalah klasifikasi dan regresi yang kompleks.

Dari ke empat jenis kernel dapat di gambarkan dengan menggunakan persamaan berikut (Primartha & Prof, 2021).

- Kernel Linear

$$K(x, x_k) = x_k^{Tx} \tag{2.1}$$

- Kernel Polynomial

$$K(x, x_k) = (x_k^{Tx} + 1)^d \quad (2.2)$$

- Kernel RBF

$$K(x, x_k) = \exp\left(-\frac{\lVert x - x_k \rVert_2^2}{\sigma^2}\right) \quad (2.3)$$

- Kernel Sigmoid

$$K(x, x_k) = \tanh(\kappa x_k^{Tx} + \theta) \quad (2.4)$$

Model SVM, yang merupakan metode klasifikasi dalam *machine learning*, dapat digambarkan dengan menggunakan persamaan berikut.

$$y(x_i) = w^T x_i + w_0 \quad (2.5)$$

Dimana y(xi) berfungsi sebagai prediksi dari tiw adalah vektor bobot (parameter model), dan xi adalah variabel data, sementara wo adalah bias. Untuk mengklasifikasikan data dalam ruang yang sesuai, model SVM bertujuan untuk memprediksi *hyperplane* dalam dimensi m. Pada dasarnya, subruang dalam bidang dua dimensi disebut sebagai *hyperplane* (Lestari et al., 2023).

2.2.5 Linear

Dalam metode *Support Vector Machine* (SVM), Data dipetakan ke dalam ruang fitur yang lebih besar melalui penggunaan kernel, yang memungkinkan SVM untuk menemukan *hyperplane* pemisah terbaik. Salah satu jenis kernel yang mudah namun sangat efektif adalah kernel linear. Fungsi kernel linear ini dapat dirumuskan dalam bentuk berikut:

$$K(x_i, x) = x_i^T x \quad (2.6)$$

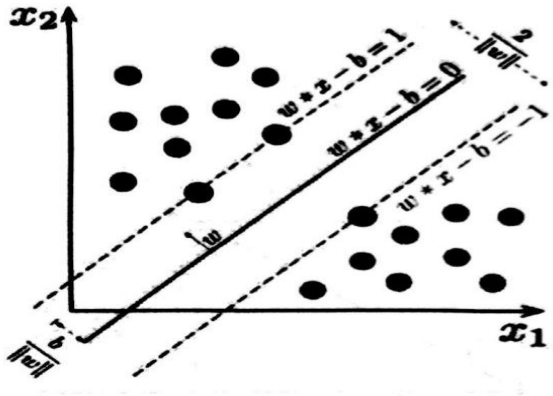
Keterangan :

a. $K(x_i, x)$: Fungsi Kernel
b. x_i : Vektor data ke – i
c. x : Vektor input
d. $x_i^T x$: Dot Product antara x_i dan x

linear SVM adalah algoritma SVM yang dapat memisahkan dua buah *class* dengan menggunakan *hyperplane* berupa garis lurus. *Hyperplane* dapat dinyatakan melalui persamaan berikut (Primartha & Prof, 2021).

$$\vec{w} \cdot \vec{x} - b = 0 \qquad (2.7)$$

Di mana $\vec{w}$ merupakan normal *vector* untuk *hyperplane*.



Gambar 2. 5 (Primartha & Prof, 2021)

Pada gambar 2.3 merupakan linear SVM yang dapat memisahkan dua buah *class*, di mana semua input *vector* yang ada pada *boundary* atau di atas *boundary* akan di kelompokkan dalam *class* 1, sedangkan semua input *vector* yang ada pada *boundary* atau di bawah *boundary* akan dikelompokkan dalam *class* -1.

2.2.6 SMOTE (*Synthetic Minority Over-sampling Technique*)

Teknik SMOTE atau *Synthetic Minority Oversampling Technique* adalah metode yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam data klasifikasi. Metode ini adalah untuk membuat data kelas yang kurang banyak menjadi lebih banyak, sehingga distribusi kelas dalam dataset menjadi lebih seimbang dengan kelas yang jumlahnya lebih banyak. Ketidakseimbangan kelas sering menyebabkan model memprediksi kelas yang jumlahnya lebih banyak dengan lebih baik, dan kinerja model akhirnya buruk dalam mendeteksi kelas yang jumlahnya sedikit. Dengan menambahkan data baru yang dibuat secara sintetis ke dalam kelas minoritas, SMOTE bisa meningkatkan kinerja model karena data yang digunakan lebih mewakili seluruh kelas. Adapun rumus dari SMOTE dapat dijelaskan sebagai berikut (Setiawan et al., 2025).

$$X_{\text{syn}} = X_i + (X_{\text{knn}} - X_i) \times \delta \qquad (2.8)$$

Keterangan :

a. X_{syn} : Data synthetic yang akan diciptakan
b. X_i : Data yang akan di replikasi
c. X_{knn} : Data yang memiliki jarak dari data X_i
d. δ : Angka acak antara 0 sampai 1

Proses penyeimbangan data menggunakan SMOTE dilakukan dengan cara mengidentifikasi kelas yang jumlah datanya paling sedikit dibandingkan kelas lainnya (kelas minoritas). Setelah itu, untuk setiap data pada kelas minoritas tersebut, dicari beberapa tetangga terdekat menggunakan k-Nearest Neighbors (k-NN). Berikutnya adalah pembuatan data sintesis yaitu dengan menghitung selisih antara data asli dan salah satu tetangga terdekatnya, kemudian mengalikannya dengan nilai acak antara 0 hingga 1, lalu menambahkannya kembali ke data asli. Selanjutnya, penambahan data sintesis, data yang telah dihasilkan ditambahkan dalam dataset sehingga jumlah data pada kelas minoritas menjadi lebih seimbang dengan kelas mayoritas (Al-Afghoni et al., 2025).

2.2.7 *Cross-Industry Standard Process for Data Mining* (CRISP-DM)

Metode CRISP-DM *(Cross Industry Standard Process for Data Mining)* adalah salah satu pendekatan yang populer dalam data mining karena dapat menghubungkan kebutuhan bisnis dengan proses pengolahan data secara terstruktur. Metode ini juga memfasilitasi pengembangan aplikasi dan menyatukan berbagai keahlian dalam bidang Teknologi Informasi. CRISP-DM mencakup enam tahapan utama, yaitu Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Penyebaran, yang saling terhubung untuk menghasilkan solusi berbasis data yang efektif (Siswipraptini, Fadiarora, et al., 2023).

Gambar 2. 6 CRISP-DM

Sumber : (Cmotions, 2023)

Adapun penjelasan tahapan dalam proses CRISP-DM pada gambar 2.4 adalah sebagai berikut:

1. *Business Understanding*

   Tahap ini merupakan langkah awal yang menitikberatkan pada pemahaman terhadap konteks bisnis atau penelitian, meliputi penentuan tujuan, perumusan permasalahan, serta identifikasi kebutuhan dalam proses *data mining*. Tujuan utama dari tahap ini adalah untuk memastikan bahwa kegiatan analisis data berjalan sejalan dengan target atau hasil yang ingin dicapai.

2. *Data Understanding*

   Pada tahap ini, dilakukan proses pengumpulan serta eksplorasi awal terhadap data yang akan dimanfaatkan dalam penelitian. Data tersebut dianalisis guna memahami karakteristik, struktur, dan kualitasnya, serta untuk memperoleh gambaran awal mengenai pola atau informasi signifikan yang mungkin terdapat di dalamnya.

3. *Data Preparation*

   Tahap ini mencakup proses pembersihan, pengolahan, dan penyiapan data sebelum digunakan dalam tahap analisis. Kegiatan yang dilakukan meliputi penyaringan, penggabungan, transformasi, serta seleksi fitur, dengan tujuan memastikan bahwa data telah berada dalam kondisi optimal dan siap untuk digunakan pada tahap pemodelan.

4. *Modeling*

   Pada tahap ini, dilakukan pemilihan metode pemodelan yang paling tepat sesuai dengan tujuan penelitian, disertai dengan proses penyesuaian parameter model guna memperoleh kinerja dan hasil yang paling optimal.

5. *Evaluation*

   Tahap evaluasi dilakukan untuk menilai sejauh mana kinerja model yang telah dikembangkan, serta memastikan bahwa model tersebut telah sesuai dengan tujuan dan kebutuhan yang ditetapkan pada tahap perencanaan awal.

2.2.8 Python

Python adalah bahasa pemrograman tingkat tinggi yang dinamis dan menggunakan sistem interpreter, yang mendukung paradigma pemrograman berorientasi objek dalam pengembangan aplikasi dan menawarkan berbagai struktur data tingkat tinggi. Selain itu, bahasa ini dapat menerjemahkan kode sumber, atau *source code*, langsung ke kode komputer saat program dijalankan. *Python* dikenal sebagai bahasa skrip yang sederhana namun kuat dan serbaguna, sehingga banyak digunakan dalam pengembangan aplikasi. Dengan sintaks yang jelas dan pengetikan yang dinamis, *Python* sangat ideal untuk pembuatan skrip maupun pengembangan

aplikasi secara cepat. Selain itu, *Python* bersifat multiguna, karena dapat diterapkan pada berbagai bidang seperti pengembangan web, aplikasi enterprise, hingga desain CAD 3D. Banyak *framework* dan *library* tersedia untuk *Python* yang sangat berguna, seperti Django untuk pengembangan web, *TensorFlow* dan *Scikit-learn* untuk *machine learning*, serta *Pandas* dan *NumPy* untuk analisis data. *Python* sangat populer dan banyak digunakan dalam industri teknologi saat ini karena kemampuan untuk terintegrasi dengan banyak sistem dan *platform*. (Suharto Agus, 2023).

2.2.9 Google Collab

Google Colaboratory, atau *Google Colab*, adalah sebuah layanan komputasi cloud yang diberikan oleh Google untuk mendukung penelitian ilmiah dan kegiatan pengembangan perangkat lunak. Colab, produk penelitian Google, memungkinkan pengguna menulis, menjalankan, dan menguji kode. *Google Colaboratory*, juga dikenal sebagai *Google Colab* adalah layanan komputasi *cloud* yang diberikan oleh Google untuk mendukung kegiatan penelitian ilmiah dan pengembangan perangkat lunak. Produk penelitian Google ini memungkinkan pengguna menulis, menjalankan, dan menguji kode. Platform ini sangat sesuai untuk kebutuhan *machine learning*, analisis data, serta keperluan pendidikan karena telah terintegrasi dengan berbagai pustaka *Python* yang umum digunakan. Colab juga mendukung penggunaan sumber daya komputasi seperti CPU, GPU, dan TPU. Selain itu, Colab mendukung kerja kolaboratif secara daring, sehingga beberapa pengguna dapat mengakses dan mengembangkan kode yang sama secara bersamaan, mirip dengan konsep kolaborasi pada *Google Docs*. Fitur tersebut menjadikan *Google Colab* sangat membantu dalam kerja tim, khususnya dalam pengembangan aplikasi dan penelitian berbasis data (Wilyani et al., 2024).

2.2.10 Confusion Matrix

Dalam tugas klasifikasi, tabel penilaian *confusion matrix* digunakan untuk mengukur seberapa baik suatu model dapat mengkategorikan data ke dalam kelas yang tepat. Tabel ini menyajikan pembandingan antara hasil prediksi model dengan label kelas yang sesungguhnya, sehingga memungkinkan kita untuk mengetahui berapa banyak data yang diklasifikasikan secara akurat dan berapa yang salah.

Dengan bantuan *confusion matrix*, berbagai ukuran evaluasi seperti akurasi, presisi, recall, serta skor f1 dapat dihitung untuk menilai tingkat kebenaran dan efektivitas model dalam membedakan antar kelas (Umam & Handoko, 2024).

Akurasi adalah metrik yang digunakan untuk melihat seberapa tepat suatu model dalam menghasilkan prediksi yang sesuai dengan kondisi sebenarnya berdasarkan data input yang diberikan, dengan rumus perhitungannya seperti pada Persamaan (2.6). Presisi adalah metrik evaluasi yang menunjukkan seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model, dengan rumus perhitungannya seperti pada Persamaan (2.7). Recall adalah metrik yang digunakan untuk melihat seberapa banyak data yang benar-benar termasuk dalam kelas positif dan berhasil dikenali dengan tepat oleh model. Nilai ini diperoleh dengan membandingkan jumlah prediksi positif yang benar terhadap seluruh data positif yang sebenarnya ada di dalam dataset, dengan rumus perhitungannya seperti pada Persamaan (2.8). F1-Score adalah ukuran evaluasi yang diperoleh dari gabungan nilai precision dan recall melalui perhitungan rata-rata harmonik. Metrik ini digunakan untuk melihat keseimbangan kinerja model, dengan rumus perhitungannya seperti pada Persamaan (2.9) (Rifa'i & Ardiyanto, 2025).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.6}$$

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{2.7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2.8}$$

$$F_{\text{1-score}} = \frac{2\times\text{Presisi}\times\text{Recall}}{\text{Presisi}+\text{Recall}} \tag{2.9}$$

Keterangan :

a. TP = *True Positive*

b. TN = *True Negative*

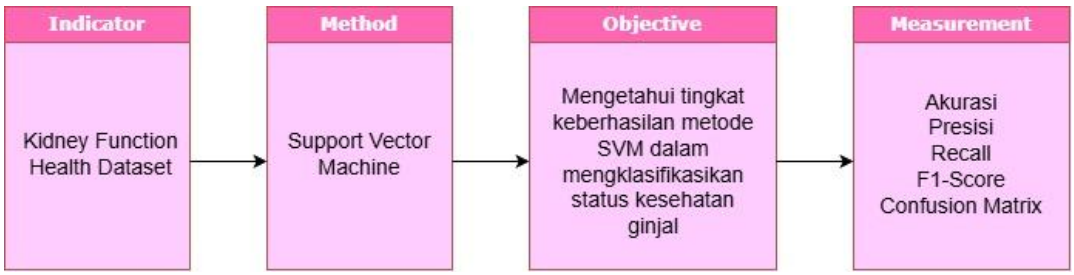
c. FP = *False Positive*

d. FN = *False Negative*

Confusion Matrix memiliki empat kategori utama yang menunjukkan hasil klasifikasi yang dilakukan oleh model. *True Positive* (TP) adalah jumlah data positif yang diprediksi benar sebagai positif dan memang benar-benar positif. *False Positive* (FP) adalah jumlah data yang sebenarnya negatif, namun diprediksi sebagai positif oleh model, yang dikenal sebagai kesalahan tipe I (Type I error). *True Negative* (TN) adalah jumlah data negatif yang diprediksi dengan benar sebagai negatif. Sedangkan *False Negative* (FN) adalah jumlah data yang sebenarnya positif, tetapi diprediksi sebagai negatif, yang disebut sebagai kesalahan tipe II (Type II error) (Siswipraptini, Haris, et al., 2023).

2.3 Kerangka Pemikiran

Untuk menghasilkan kesimpulan tentang apa yang sedang dilakukan dalam penelitian, kerangka pemikiran dibuat untuk menggambarkan awal dan akhir proses penelitian. Kerangka pemikiran untuk penelitian ini adalah sebagai berikut:



Gambar 2. 7 Kerangka Pemikiran

1. *Indicator*

   Indikator adalah ukuran atau tolak ukur yang digunakan dalam penelitian untuk menilai suatu variabel tertentu. Data masukan yang digunakan indikator ini akan diolah dan dianalisis lebih lanjut untuk mencapai hasil penelitian yang diinginkan. Pada tahap ini, penulis menggunakan data *Kidney Function Health Dataset* sebagai indikator utama yang diperoleh dari web yaitu Kaggle.com. Dataset ini berisi sejumlah atribut atau variabel yang merepresentasikan kondisi fungsi ginjal.

2. *Method*

   Tahap *Method* merupakan bagian yang menjelaskan pendekatan dan metode yang diterapkan dalam menyelesaikan penelitian ini. Data yang diperoleh dari situs Kaggle akan diolah dan dianalisis menggunakan metode *Support Vector Machine* (SVM).

3. *Objective*

   Objective merupakan suatu tujuan dari metode yang diterapkan adalah untuk mendapatkan hasil tingkat keberhasilan metode *Support Vector Machine* (SVM) dalam mengklasifikasikan status kesehatan ginjal.

4. *Measurement*

   Measurement adalah pengukuran tingkat akurasi dari hasil penelitian berdasarkan metode yang di terapkan. Pada penelitian ini untuk mengukur tingkat akurasi menggunakan metrik, yaitu :

   a. Akurasi: mengukur tingkat ketepatan model dalam melakukan klasifikasi secara keseluruhan.

   b. Presisi: untuk mengevaluasi ketepatan prediksi kelas positif.

   c. *Recall* : mengevaluasi kemampuan model untuk menemukan semua informasi positif.

   d. *F1- Score* : merupakan harmoni rata-rata antara recall dan presisi.

   e. *Confusion Matrix*: digunakan untuk menentukan apakah hasil klasifikasi setiap kelas benar atau salah.