

Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review

Ria Suci Nurhalizah^{*1}, Rian Ardianto², Purwono³

¹Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa, Indonesia

^{2,3}Informatika, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa, Indonesia

Email: ¹riascnr02@gmail.com, ²rianardianto@uhb.ac.id, ³purwono@uhb.ac.id

Abstrak

Artikel ini menyajikan tinjauan sistematis mengenai dua paradigma utama dalam *Machine Learning* yaitu *Supervised Learning* dan *Unsupervised Learning*, dengan tujuan memberikan pemahaman mendalam tentang perbedaan, serta kelebihan dan kekurangan masing-masing metode. Penelitian ini menerapkan metode *Literature Review* (SLR) berdasarkan pedoman PRISMA untuk menganalisis studi-studi relevan yang dipublikasikan dalam lima tahun terakhir. Dari total 540 artikel yang diperoleh, 10 artikel dipilih untuk ditelaah lebih lanjut, terdiri dari lima mengenai *Supervised Learning* dan lima mengenai *Unsupervised Learning*. Hasil analisis menunjukkan bahwa *Supervised Learning* menggunakan data berlabel Systematic untuk prediksi dan klasifikasi dengan algoritma seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine* (SVM), yang umumnya menghasilkan akurasi tinggi. Di sisi lain, *Unsupervised Learning* yang tidak membutuhkan data berlabel, fokus pada eksplorasi data dan pengelompokan menggunakan algoritma seperti *K-Means*, *Artificial Neural Network* (ANN), dan *Gaussian Mixture Model* (GMM), menawarkan fleksibilitas yang lebih besar meski dengan akurasi yang biasanya lebih rendah. Kedua pendekatan ini memiliki kelebihan dan keterbatasan masing-masing, dan pemilihan metode yang tepat harus mempertimbangkan tujuan aplikasi, ketersediaan data, dan karakteristik data.

Kata kunci: *Machine Learning, Supervised Learning, Unsupervised Learning*

Abstract

This article presents a systematic review of the two main paradigms in *Machine Learning*, namely *Supervised Learning* and *Unsupervised Learning*, with the aim of providing an in-depth understanding of the differences, as well as the advantages and disadvantages of each method. This research applies the *Literature Review* (SLR) method based on PRISMA guidelines to analyze relevant studies published in the last five years. From a total of 540 articles obtained, 10 articles were selected for further review, consisting of five on *Supervised Learning* and five on *Unsupervised Learning*. The analysis shows that *Supervised Learning* uses Systematic labeled data for prediction and classification with algorithms such as *K-Nearest Neighbor* (KNN), *Naïve Bayes*, *Decision Tree*, and *Support Vector Machine* (SVM), which generally produce high accuracy. *Unsupervised Learning*, on the other hand, which does not require labeled data, focuses on data exploration and clustering using algorithms such as *K-Means*, *Artificial Neural Network* (ANN), and *Gaussian Mixture Model* (GMM), offering greater flexibility albeit with typically lower accuracy. Both approaches have their own advantages and limitations, and the selection of an appropriate method should consider the application goals, data availability, and data characteristics.

Keywords: *Machine Learning, Supervised Learning, Unsupervised Learning*

1. PENDAHULUAN

Machine Learning adalah komponen utama dalam bidang kecerdasan buatan yang digunakan untuk menyelesaikan berbagai permasalahan [1]. Ini merupakan implementasi dari kecerdasan buatan yang fokus pada pengembangan sistem yang dapat belajar sendiri tanpa perlu diprogram secara berulang [2]. *Machine Learning* adalah bidang penelitian ilmiah yang mempelajari algoritma dan model statistik yang digunakan oleh sistem komputer untuk menyelesaikan tugas-tugas tertentu tanpa instruksi eksplisit, melainkan dengan memanfaatkan pola dan inferensi sebagai penggantinya [3]. Penggunaan *Machine*

Learning bertujuan untuk mengajarkan mesin agar dapat memproses data secara lebih efisien [4]. Melalui pengidentifikasian pola dalam dataset, *Machine Learning* mampu meramalkan dan memahami karakteristik dari objek yang tidak dikenal [5].

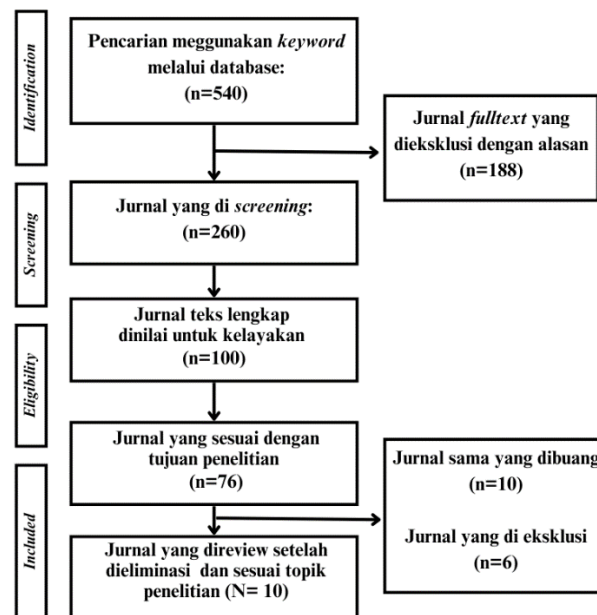
Manfaat menggunakan *machine learning* dalam peninjauan sistematis adalah dapat mengurangi jumlah pekerjaan manual yang diperlukan dalam proses tersebut [6]. *Machine Learning* menggunakan algoritma yang mampu mempelajari pola dan hubungan yang kompleks dalam data, daripada mengandalkan pendekatan berbasis aturan. Ini memungkinkan pengguna untuk melakukan keputusan dengan tingkat akurasi yang lebih tinggi [7]. Di samping itu, ada kekurangan seperti rentannya sistem yang terintegrasi dalam *Machine Learning* terhadap penyusupan dan manipulasi melalui serangan yang bersifat merugikan. Dalam serangan semacam itu, pihak yang menyerang dapat memanfaatkan kerentanan dalam sistem untuk salah mengklasifikasikan aktivitas yang sebenarnya tidak berbahaya sebagai aktivitas berbahaya [8]. Ketidakseimbangan dalam data merupakan tantangan bagi algoritma *Machine Learning* karena mereka biasanya cenderung memprediksi dengan benar kelas mayoritas, namun sering kali melakukan kesalahan dalam mengidentifikasi kelas minoritas [9].

Ada tiga paradigma utama yang digunakan untuk menggali pola dan pengetahuan dari data: *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. Dalam *Supervised Learning*, model pembelajaran menggunakan data yang sudah diberi label sebagai dasar untuk prediksi atau klasifikasi [10]. Sedangkan dalam *Unsupervised Learning* model harus menemukan pola dari data tanpa bantuan label eksternal [11]. Meskipun tujuan dari kedua paradigma ini adalah untuk menghasilkan model yang mampu mengenali pola dalam data, perbedaan mendasar dalam pendekatan mereka memiliki dampak yang signifikan dalam aplikasi praktis dan hasil yang diharapkan. Sedangkan dalam *Reinforcement Learning* adalah sebuah algoritma yang mengakumulasi informasi (dalam bentuk sinyal penguatan) untuk memilih langkah-langkah yang menuju hasil yang diinginkan dengan tingkat optimal [12]. Sebuah agen yang melakukan tindakan dan memperoleh pembelajaran dari hasil tindakan tersebut [13]. Meskipun tujuan dari ketiga paradigma ini adalah untuk menghasilkan model yang mampu mengenali pola dalam data, perbedaan mendasar dalam pendekatan mereka memiliki dampak yang signifikan dalam aplikasi praktis dan hasil yang diharapkan.

Tinjauan literatur sistematis ini hanya bertujuan untuk menyajikan analisis mendalam tentang perbedaan antara *Supervised Learning* dan *Unsupervised Learning*, dengan mempertimbangkan berbagai studi dan publikasi terkait. Dengan memahami perbedaan yang lebih dalam antara *Supervised Learning* dan *Unsupervised Learning*, kita dapat menginformasikan pemilihan model yang tepat untuk berbagai aplikasi, serta memberikan pandangan yang lebih luas tentang kemungkinan penelitian masa depan di bidang pembelajaran mesin dan analisis data.

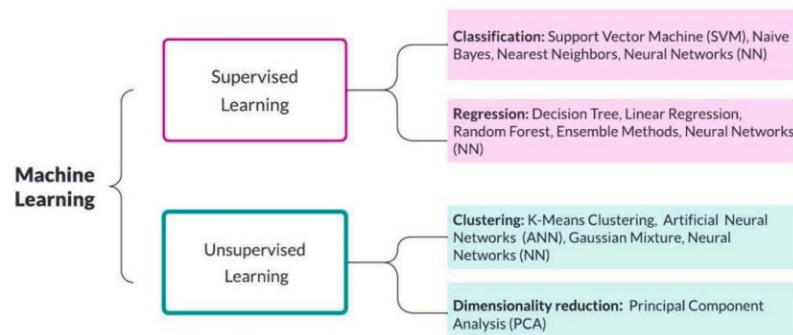
2. METODE PENELITIAN

Penelitian ini menerapkan metode *Systematic Literature Review* (SLR) dengan mengacu pada pedoman *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA). Langkah-langkah penelitian dijelaskan melalui flowchart pada Gambar 1. Proses pencarian literatur dilakukan melalui berbagai search engine dan database akademis terkemuka seperti *Google Scholar*, *Science Direct*, *IEEE Xplore*, *ScienceDirect* serta jurnal ilmiah lainnya dengan mengumpulkan informasi dari penelitian terdahulu menggunakan kata kunci pencarian yang terfokus pada *Machine Learning*, *Supervised Learning* dan *Unsupervised Learning*. Pencarian ini mencakup literatur yang dipublikasikan dalam rentang waktu kurang lebih lima tahun terakhir untuk memastikan relevansi dan kebaruan informasi yang diperoleh. Dari pencarian awal menggunakan kata kunci tersebut, diperoleh 540 artikel. Setelah data literatur terkumpul, peneliti melakukan seleksi yang menyisakan 260 artikel untuk disaring lebih lanjut. Dari 260 artikel tersebut, 100 dipilih untuk dinilai kelayakannya berdasarkan teks lengkap dan 76 artikel dinyatakan sesuai dengan tujuan penelitian. Pada tahap akhir 16 artikel dikeluarkan karena alasan seperti duplikasi atau ketidaksesuaian dengan kriteria akhir, sehingga hanya 10 jurnal yang dipilih untuk di review dan dianalisis dalam penelitian ini.



Gambar 1. Flowchart PRISMA

3. HASIL DAN PEMBAHASAN



Gambar 2. Algoritma Mechine Learning yang banyak digunakan [24]

Pada dasarnya, *machine learning* terbagi tiga, yaitu *supervised learning*, *unsupervised learning*. *Supervised learning* merupakan metode dalam *machine learning* yang memanfaatkan data yang telah diberi label atau dataset yang sudah diketahui oleh pembuatnya. Data yang telah dilabelkan tersebut berfungsi untuk melatih algoritma dengan bimbingan, sehingga algoritma dapat mengklasifikasikan atau memprediksi suatu kasus dengan tingkat akurasi yang baik. Sedangkan *Unsupervised learning* merupakan metode dalam *machine learning* yang mengaplikasikan algoritma untuk menganalisis dan menemukan pola dalam data tanpa intervensi atau bantuan manusia [25]. Dalam konteks ini, tidak diberikan informasi tentang hasil yang diharapkan dari suatu input terhadap algoritma, dan algoritma bertugas menemukan pola yang mungkin terdapat dalam dataset tersebut.

Berdasarkan penelusuran studi pada berbagai database, terpilih sepuluh studi yang dapat dijadikan referensi untuk penelitian yang telah direncanakan yang terdiri dari lima pembahasan tentang *Supervised Learning* dan lima membahas tentang *Unsupervised Learning*. Berikut adalah ringkasan hasil evaluasi:

3.1. Supervised Learning

Tabel 1. Studi literatur mengenai *Supervised Learning*

No	Judul	Author	Metode	Tahun	Hasil
1	<i>Usability</i> Algoritma <i>Supervised Learning</i> Untuk Prediksi Kelulusan Mahasiswa Pada Sistem Bimbingan Akademik [14]	Triase, Sriani, Khairuna	<i>K-Nearest Neighbor</i>	2022	Penggunaan algoritma supervised learning dengan metode <i>K-Nearest Neighbor</i> untuk memprediksi kelulusan tepat waktu menunjukkan akurasi 96% dalam memprediksi mahasiswa yang tidak lulus. Dari hasil tersebut, perempuan (11) lebih banyak lulus tepat waktu dibandingkan laki-laki (5). Secara keseluruhan, kelulusan didominasi oleh perempuan (154) dibandingkan laki-laki (147). Waktu studi terbanyak adalah semester 9 dengan jumlah 170 mahasiswa.
2	Penerapan Algoritma <i>Supervised Learning</i> untuk Klasifikasi Data Music Listening [15]	Eri Mardiani	Menggunakan lima algoritma data mining sesuai dengan persyaratan Teknik Klasifikasi	2023	Setelah melakukan analisis dan pengujian data untuk memprediksi preferensi musik menggunakan algoritma <i>K-Nearest</i> , <i>Naive Bayes</i> , <i>Decision Tree</i> , <i>ensemble method</i> , dan <i>linear regression</i> , ditemukan bahwa <i>Naive Bayes</i> memiliki akurasi tertinggi dengan rata-rata 87%. Diikuti oleh <i>K-Nearest</i> dengan 76,1%, <i>Decision Tree</i> dengan 56,7%, dan <i>Linear Regression</i> dengan 6,3%.
3	Penerapan Algoritma <i>Supervised Learning</i> untuk Klasifikasi Program Keluarga Harapan [16]	Muhammad Syarif Hartawan	Kajian literatur dan pengumpulan data, pra-pemrosesan dan pengolahan data, perbandingan algoritma, serta pengambilan keputusan.	2023	Berdasarkan analisis, kesimpulan utama adalah pemilihan metode terbaik didasarkan pada tingkat akurasi tertinggi. Algoritma <i>Naive Bayes Classifier</i> memiliki akurasi tertinggi, yaitu 93,2%, dibandingkan dengan <i>K-Nearest Neighbor</i> sebesar 82,3% dan <i>Probabilistic Neural Network</i> sebesar 89,7%. Oleh karena itu dalam kasus ini, algoritma <i>Naive Bayes Classifier</i> merupakan pilihan terbaik untuk digunakan dalam pemodelan pengambilan keputusan.
4	Klasifikasi Penyakit Gagal Jantung Menggunakan <i>Supervised Learning</i> [17]	Ena Tasia	Proses penelitian ini terdiri dari lima tahap yang mencakup pengumpulan data, pra-pemrosesan data, klasterisasi menggunakan algoritma, evaluasi validitas klaster, dan analisis serta penyimpulan hasil dari algoritma terbaik.	2023	Dari hasil penelitian, dapat disimpulkan bahwa penyakit gagal jantung memiliki risiko kematian yang tinggi. Oleh karena itu, dilakukan klasifikasi menggunakan tiga algoritma: C4.5, <i>Support Vector Machine</i> , dan <i>Naive Bayes Classifier</i> . Hasil klasifikasi menunjukkan akurasi masing-masing algoritma sebesar 80%, 68%, dan 75%. Algoritma C4.5 memiliki akurasi tertinggi, yaitu 80%, menunjukkan keunggulannya dalam memprediksi penyakit gagal jantung dibandingkan <i>Support Vector Machine</i> dan <i>Naive Bayes Classifier</i> .
5	Analisis Sentimen Ulasan Pengunjung Terhadap Tempat Wisata Religi Walisongo	Pandu Rizki Maulidiah	Untuk mempersiapkan model, ulasan yang di-scraping akan diberi label positif, negatif, dan netral	2023	Hasil penelitian menunjukkan bahwa, dibandingkan dengan model SVM-RBF lainnya, model ini memiliki akurasi tertinggi sebesar 87,12% dan skor f1 tertinggi untuk kategori negatif (89%), netral (68%), dan positif (91%).

Menggunakan	secara	manual.
Metode	Menggunakan	
<i>Supervised</i>	algoritma	seperti
<i>Learning</i> [18]	<i>Decision Tree</i>	
	<i>Classifier, K-Nearest</i>	
	<i>Neighbor,</i>	
	<i>Multinomial Naïve</i>	
	<i>Bayes, dan Support</i>	
	<i>Vector Machine</i>	
	kernel RBF, data yang	
	telah diproses akan	
	digunakan dalam	
	model. Laporan	
	klasifikasi dan	
	<i>confusion matrix</i> akan	
	digunakan untuk	
	menguji kinerja	
	model.	

Supervised Machine Learning merupakan sub-bidang dari *Machine Learning* yang umumnya memanfaatkan ahli di bidangnya untuk mengawasi skema pembelajaran yang dibutuhkan. *Supervised Machine Learning* mengarah pada tugas yang memetakan *input* ke *output* yang telah ditentukan. *Supervised Machine Learning* sering digunakan dalam masalah klasifikasi karena tujuannya adalah membuat komputer memahami kerangka kerja deskriptif yang telah disiapkan [26]. *Supervised Machine Learning* adalah bidang yang melintasi berbagai disiplin ilmu seperti ilmu komputer, statistik, ilmu kognitif, teknik, teori optimisasi, serta berbagai cabang matematika dan sains lainnya [27]. *Supervised Learning* digunakan untuk memprediksi hasil tertentu dengan tujuan melakukan estimasi atau klasifikasi. [28]. Beberapa algoritma umum yang digunakan dalam klasifikasi *supervised learning* mencakup [29] [30]:

a. *K-Nearest Neighbor*

Metode sederhana dan nonparametrik untuk klasifikasi adalah klasifikasi *K-Nearest Neighbor* [31]. Algoritma *K-Nearest Neighbor* adalah metode pembelajaran yang diawasi di mana kategori dari sebuah instance pertanyaan baru ditentukan berdasarkan mayoritas kategori dari tetangga terdekat K [32]. *K-Nearest Neighbor* juga digunakan untuk mengklasifikasikan objek dengan membandingkan jaraknya dengan data pelatihan [33].

b. *Naïve Bayes*

Metode *Naïve Bayes* merupakan teknik yang digunakan untuk memperkirakan probabilitas [34] [35]. Klasifier *Naïve Bayes* adalah salah satu pendekatan yang efektif untuk membuat keputusan untuk meningkatkan hasil dalam masalah klasifikasi [36]. Algoritma *Naïve Bayes* berusaha untuk menemukan pola atau pengetahuan yang serupa dalam kelompok atau kelas tertentu [37].

c. *Decision Tree*

Decision tree adalah model prediksi yang menggunakan struktur pohon atau hirarki untuk mengambil keputusan [38]. Setiap pohon memiliki cabang yang menunjukkan suatu tugas yang harus dilakukan untuk melanjutkan ke cabang berikutnya dan akhirnya mencapai daun yang menunjukkan hasil prediksi [39]. Pohon keputusan juga dapat mengubah informasi menjadi struktur pohon keputusan yang memvisualisasikan aturan dalam bentuk yang terstruktur [40]. Ketika data dari tabel diubah menjadi model pohon, aturan disederhanakan dibuat [41].

d. *Support Vector Machine*

Algoritma *Support Vector Machine* menggunakan pemetaan nonlinier untuk mengubah data penelitian ke dimensi yang lebih besar [42]. *Support Vector Machine* mampu mengelola data dalam dimensi yang tinggi, sehingga meningkatkan tingkat akurasi prediksi [43]. Prinsip utama *Support Vector Machine* dalam mengklasifikasikan data adalah menemukan hyperplane terbaik yang optimal memisahkan dua kelas yang sudah ditentukan [44] [45].

3.2. Unsupervised Learning

Tabel 2. Studi literatur tentang *Unsupervised Learning*

No	Judul	Author	Metode	Tahun	Hasil
1	<i>Unsupervised Learning of Compositional Energy Concepts</i> [19]	Yilun Du, Shuang Li, Yash Sharma, Joshua B. Tenenbaum, Igor Mordatch	Metode yang digunakan dalam adalah COMET, yang merupakan sebuah pendekatan untuk menemukan dan merepresentasikan konsep sebagai fungsi energi terpisah	2021	Artikel tersebut membahas pengembangan metode baru yang disebut COMET untuk pembelajaran tanpa pengawasan dalam gambar. COMET dapat mengidentifikasi dan memisahkan faktor-faktor variasi dalam gambar, seperti posisi kamera dan segmentasi objek, secara fleksibel dan efisien. Dibandingkan dengan model lain seperti β -VAE dan InfoGAN, COMET terbukti lebih baik dalam tugas disentanglement. Analisis menunjukkan bahwa penggunaan fungsi energi dalam COMET lebih ekspresif dan efisien secara komputasi daripada masker <i>segmentasi</i> . Peneliti juga menekankan pentingnya dataset yang seimbang dan adil saat menggunakan COMET.
2	Implementasi <i>unsupervised learning</i> pada Nilai Jasmani Kesamaptaan Sekolah Polisi Negara dengan Metode <i>Clustering Analysis</i> [20]	Zaenal Muttaqin, Donny Fernando, Selvia Sulastriani	Metode <i>Single Linkage, Average Linkage</i> dan <i>Complete Linkage</i> .	2023	Dengan menggunakan metode perhitungan jarak Manhattan dengan penghubung satu, dihasilkan 17 kluster yang beragam. Sebaliknya, dengan menggabungkan penghubung rata-rata dengan perhitungan jarak Manhattan, dihasilkan 24 kluster. Analisis akhir menunjukkan bahwa metode penghubung lengkap dengan perhitungan jarak geometri juga menghasilkan 24 kluster.
3	Implementasi Metode <i>Unsupervised Learning</i> Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP [21]	Desta Yolanda, Mohammad Hafiz Hersyah, Eno Marozi	Metode penelitian eksperimental (<i>experimental research</i>)	2021	Pengujian dan analisis menunjukkan bahwa sistem dapat membedakan wajah yang dikenali dan tidak dikenali dengan metode pembelajaran yang tidak diawasi, berdasarkan gagasan kelompok gambar. Waktu rata-rata 2,15 detik diperlukan untuk pemrosesan gambar wajah hingga hasil yang ditampilkan di antarmuka pengguna. Dengan ukuran file rata-rata sekitar 2,28 MB selama <i>streaming</i> 24 jam, sistem ini juga menghemat hingga 99% penyimpanan. Secara keseluruhan, sistem ini berfungsi dengan baik dan memenuhi harapan.
4	Perbandingan Teknik <i>Unsupervised Learning</i> untuk Pengelompokan Data Jumlah Desa Di Indonesia [22]	Akhiril Anwar Harahap, Muhammad Raiha, Nailul Amani, Putri Risma Andini	Pengumpulan dan <i>pre-processing</i> data, klusterisasi data, pengukuran validasi cluster pada algoritma, dan analisis hasilnya.	2023	Dengan menggunakan DBI, setiap algoritma diuji dengan percobaan $K=3$ hingga $K=10$. Hasilnya menunjukkan <i>K-Means</i> 0,451, <i>K-Medoids</i> 0,638, dan <i>Fuzzy-C-Means</i> 0,491. <i>K-Means</i> terbukti lebih baik dalam mengelompokkan data desa dengan fasilitas sekolah berdasarkan provinsi Indonesia.
5	<i>Unsupervised Learning of Dense Visual</i>	Pedro O. Pinheiro, Amjad	Metode momentum <i>contrast mechanism</i> untuk mempelajari	2020	Metode VADeR untuk pembelajaran representasi visual yang padat melalui pembelajaran kontras tingkat piksel

Representations [23]	Almahairi, Ryan Y. Benmalek, Florian Golemo, Aaron Courville	representasi visual yang padat dari data tanpa label	berhasil mengungguli baseline kuat dalam berbagai tugas prediksi terstruktur. Pembelajaran representasi padat tanpa pengawasan lebih efisien untuk tugas <i>downstream</i> dibandingkan dengan representasi global, bahkan mengungguli pretraining terawasi ImageNet dalam beberapa tugas . Metode ini menyoroti pentingnya pembelajaran representasi padat dalam memahami gambar dengan data berlabel terbatas
----------------------	--	--	---

Unsupervised machine learning adalah metode yang digunakan tanpa memerlukan data pelatihan [46]. Dalam *Unsupervised learning*, pengamatan atau data digunakan tanpa label, kelas, atau keputusan yang telah ditentukan [47][48]. Metode pembelajaran tanpa pengawasan sangat umum digunakan untuk berbagai tugas statistik yang penting, dengan aplikasi yang luas dalam *sains* dan industri. [49]. *Unsupervised machine learning* semakin umum dalam memprediksi cacat perangkat lunak. Pendekatan ini dapat bermanfaat bagi praktisi perangkat lunak karena mengurangi ketergantungan pada data pelatihan yang telah dilabeli [50]. Pada *Unsupervised learning* menilai hasil dari metode pembelajaran tanpa pengawasan atau menganalisis data menjadi sulit karena tidak ada mekanisme yang secara umum diterima untuk melakukan validasi silang atau memverifikasi hasil pada set data yang independen [51]. Beberapa algoritma yang digunakan dalam *clustering unsupervised learning* seperti :

- a. *K-Means*
Algoritma *K-Means* merupakan salah satu algoritma dalam data mining yang menggunakan teknik *clustering* berdasarkan pembagian jarak [52]. *K-Means* digunakan untuk mengelompokkan data berdasarkan kemiripan (*similarity*) [53]. Metode *K-means* adalah teknik yang efisien untuk mengelompokkan data dalam jumlah besar dengan waktu komputasi yang cepat.
- b. *Artificial Neural Network*
Artificial neural network didasarkan pada struktur saraf manusia dan memungkinkan fungsi penjumlahan, aktivasi, dan tujuan [54] [55] [56]. Arsitektur jaringan buatan terdiri dari tiga lapisan *node* yang saling terhubung: lapisan *input*, lapisan tersembunyi, dan lapisan *output* [57].
- c. *Gaussian Mixture*
Gaussian Mixture Model adalah model yang menggabungkan distribusi *Gaussian* [58] [59]. *Gaussian Mixture Model* adalah model *probabilitas* yang mengasumsikan bahwa data mengikuti karakteristik dari beberapa distribusi *Gaussian* dengan parameter yang tidak diketahui [60]. *Gaussian Mixture Model* merupakan model statistik yang mengestimasi probabilitas dan bobot dari setiap distribusi *Gaussian*, sehingga dapat digunakan dengan baik baik saat parameter diketahui maupun tidak.

Tabel 3. Perbedaan *Supervised* dan *Unsupervised Machine Learning* [61]

Kriteria	Supervised Learning	Unsupervised Learning
Definisi	Metode pembelajaran mesin yang menggunakan data berlabel untuk melatih model agar dapat memprediksi atau mengklasifikasi data baru dengan akurasi tinggi.	Metode pembelajaran mesin yang tidak menggunakan data berlabel, berfokus pada identifikasi pola atau struktur yang tidak diketahui dalam dataset.
Algoritma Umum	<i>K-Nearest Neighbor</i> , <i>Naïve Bayes</i> , <i>Decision Tree</i> , <i>Support Vector Machine</i> .	<i>K-Means Clustering</i> , <i>Artificial Neural Network</i> , <i>Gaussian Mixture Model</i> .
Penggunaan Data	Memerlukan data berlabel untuk melatih model, sehingga model dapat mengasosiasikan input dengan output yang tepat.	Tidak memerlukan data berlabel, model mengidentifikasi pola atau struktur berdasarkan karakteristik data itu sendiri.

Tujuan Utama	Prediksi atau klasifikasi data baru dengan meminimalkan kesalahan atau meningkatkan akurasi berdasarkan data pelatihan.	Eksplorasi dan pengelompokan data, menemukan hubungan tersembunyi atau pola dalam data tanpa panduan eksplisit.
Keuntungan	<ul style="list-style-type: none">a. Hasil yang lebih akurat dan dapat diandalkan jika data berlabel tersedia dan representatif.b. Sangat cocok untuk masalah klasifikasi dan regresi.	<ul style="list-style-type: none">a. Dapat digunakan untuk menemukan struktur atau pola baru dalam data.b. Lebih fleksibel untuk diterapkan pada berbagai jenis data.
Keterbatasan	<ul style="list-style-type: none">a. Memerlukan sejumlah besar data berlabel yang mungkin sulit diperoleh.b. Rentan terhadap overfitting jika model terlalu kompleks atau data pelatihan tidak mencukupi.	<ul style="list-style-type: none">a. Kurang akurat dibandingkan dengan supervised learning karena tidak ada data berlabel.b. Hasilnya mungkin lebih sulit untuk ditafsirkan dan divalidasi.
Hasil dari studi Literatur	<ul style="list-style-type: none">a. <i>K-Nearest Neighbor</i>, <i>Naïve Bayes</i>, dan <i>Support Vector Machine</i> sering menunjukkan akurasi tinggi dalam berbagai aplikasi.b. <i>Decision Tree</i> efektif dalam memvisualisasikan aturan dan memisahkan kelas dengan baik.	<ul style="list-style-type: none">a. <i>K-Means</i> efektif dalam mengelompokkan data menjadi cluster berdasarkan kesamaan.b. <i>Gaussian Mixture Model</i> dan <i>Artificial Neural Network</i> sering digunakan untuk mengidentifikasi pola kompleks tanpa panduan eksplisit.

4. KESIMPULAN

Berdasarkan diskusi sebelumnya, *Supervised Learning* menggunakan data berlabel dan cocok untuk tugas-tugas prediksi atau klasifikasi yang memerlukan akurasi tinggi, dengan algoritma seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine* (SVM) yang efektif dalam berbagai aplikasi. Sebaliknya, *Unsupervised Learning* tidak membutuhkan data berlabel dan berfokus pada eksplorasi serta pengelompokan data untuk menemukan pola atau struktur yang tidak diketahui sebelumnya, menggunakan algoritma seperti *K-Means Clustering*, *Artificial Neural Network* (ANN), dan *Gaussian Mixture Model* (GMM). Meskipun kurang akurat dibandingkan dengan *Supervised Learning*, metode ini lebih fleksibel dalam menangani berbagai jenis data. Pemilihan pendekatan yang tepat harus mempertimbangkan tujuan penelitian, ketersediaan data, dan karakteristik data yang ada, dengan kedua metode memiliki peran penting dalam pengembangan teknologi *Machine Learning* yang efektif dan efisien.

DAFTAR PUSTAKA

[1] J. Homepage, A. Roihan, P. Abas Sunarya, and A. S. Rafika, "IJCIT (Indonesian Journal on Computer and Information Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," 2019.

[2] C. Chazar and B. E. Widhiaputra, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 12, pp. 67–80, 2020.

[3] Rizki Rino Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *Jurnal MATRIK*, vol. 19, pp. 302–311, May 2020.

[4] B. Mahesh, "Machine Learning Algorithms-A Review," *International Journal of Science and Research*, 2018, doi: 10.21275/ART20203995.

[5] Mutammimul Ula, Ananda Faridhatul Ulva, and Mauliza, "Implementasi Machine Learning dengan Model Case Based Reasoning dalam Mendagnosa Gizi Buruk pada Anak," *Jurnal Informatika Kaputama (JIK)*, vol. 5, pp. 333–339, Jul. 2021.

[6] B. G. Pijls, "Machine Learning assisted systematic reviewing in orthopaedics," *J Orthop*, vol. 48, pp. 103–106, Feb. 2024, doi: 10.1016/j.jor.2023.11.051.

- [7] S. Jahandideh, G. Ozavci, B. W. Sahle, A. Z. Kouzani, F. Magrabi, and T. Bucknall, "Evaluation of machine learning-based models for prediction of clinical deterioration: A systematic literature review," *International Journal of Medical Informatics*, vol. 175. Elsevier Ireland Ltd, Jul. 01, 2023. doi: 10.1016/j.ijmedinf.2023.105084.
- [8] O. Alshaikh, S. Parkinson, and S. Khan, "Exploring Perceptions of Decision-Makers and Specialists in Defensive Machine Learning Cybersecurity Applications: The Need for a Standardised Approach," *Comput Secur*, p. 103694, Dec. 2023, doi: 10.1016/j.cose.2023.103694.
- [9] A. X. Wang, S. S. Chukova, and B. P. Nguyen, "Synthetic minority oversampling using edited displacement-based k-nearest neighbors," *Appl Soft Comput*, vol. 148, Nov. 2023, doi: 10.1016/j.asoc.2023.110895.
- [10] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," *Buletin Ilmiah Ilmu Komputer dan Multimedia*, pp. 216–219, 2022.
- [11] G. Montavon, J. Kauffmann, W. Samek, and K. R. Müller, "Explaining the Predictions of Unsupervised Learning Models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 117–138. doi: 10.1007/978-3-031-04083-2_7.
- [12] M. Riziq sirfatullah Alfarizi, M. Zidan Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, "PENGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING," 2023.
- [13] B. Aryo Dharmawan, "Analisis dan implementasi Sistem Trading Otomatis Bitcoin dengan Menggunakan Recurrent Reinforcement Learning."
- [14] Triase, Sriani, and Khairuna, "Usability Algoritma Supervised Learning Untuk Prediksi Kelulusan Mahasiswa pada Sistem Bimbingan Akademik," *Indonesian Journal of Computer Science*, vol. 11, pp. 1005–1022, 2022.
- [15] E. Mardiani *et al.*, "Penerapan Algoritma Supervised Learning untuk Klasifikasi Data Music Listening," *donesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 115–124, 2023.
- [16] M. S. Hartawan, Moh. Erkamim, S. R. Yahya, N. C. Santi, Legito, and Sepriano, "Penerapan Algoritma Supervised Learning untuk Klasifikasi Program Keluarga Harapan," *ndonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 83–91, 2023.
- [17] E. Tasia, R. Z. I. Z. Ismail, S. K. P. Loka, Y. Ikhsani, and R. Ocviani, "Metode Klasifikasi Supervised Learning pada Penyakit Gagal Jantung," *Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 1–7, 2023.
- [18] P. R. Maulidiah, A. A. Arifiyanti, and D. S. Y. Kartika, "Analisis Sentimen Ulasan Pengunjung Terhadap Tempat Wisata Religi Walisongo Menggunakan Metode Supervised Learning," *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, vol. 3, pp. 57–64, 2023.
- [19] Y. Du, S. Li, Y. Sharma, J. B. Tenenbaum, I. Mordatch, and G. Brain, "Unsupervised Learning of Compositional Energy Concepts," 2021. [Online]. Available: <https://energy-based-model.github.io/comet/>
- [20] Zaenal Muttaqi, Donny Fernando, and Selvia Sulastriani, "Implementasi Unsupervised Learning pada Nilai Jasmani Kesamaptaaan Sekolah Polisi Negara dengan Metode Clustering Analysis," *Jurnal PROSISKO*, vol. 10, pp. 18–23, Mar. 2023.
- [21] Desta Yolanda, Mohammad Hafiz Hersyah, and Eno Marozi, "Implementasi Metode Unsupervised Learning Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1099–1105, Dec. 2021, doi: 10.29207/resti.v5i6.3552.

- [22] A. A. Harahap, M. Raihan, N. Amani, and R. Andini, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Comparison of Unsupervised Learning Techniques for Clustering Data on the Number of Villages in Indonesia Perbandingan Teknik Unsupervised Learning untuk Pengelompokan Data Jumlah Desa Di Indonesia." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [23] P. O. Pinheiro, A. Almahairi, R. Y. Benmalek, F. Golemo, and A. Courville, "Unsupervised Learning of Dense Visual Representations."
- [24] G. Kantayeva, J. Lima, and A. I. Pereira, "Application of machine learning in dementia diagnosis: A systematic literature review," *Heliyon*, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21626.
- [25] T. Suhendra and M. Cs, "MAKALAH PEMBELAJARAN MESIN (MACHINE LEARNING)," 2021.
- [26] S. H. Shetty, S. Shetty, C. Singh, and A. Rao, "Supervised Machine Learning: Algorithms and Applications," 2022.
- [27] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, pp. 47–58, Apr. 2021.
- [28] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer," *Behav Ther*, vol. 51, no. 5, pp. 675–687, Sep. 2020, doi: 10.1016/j.beth.2020.05.002.
- [29] J. A. Nurcahyo and T. B. Sasongko, "Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras," *Indonesian Journal of Computer Science*, vol. 12, pp. 1351–1365, 2023.
- [30] A. Gading Pertiwi and U. Pujianto, "Metode-metode data mining untuk penyelesaian masalah kehamilan dan persalinan," *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, vol. 2, no. 1, p. 11, Apr. 2020.
- [31] P. Putra, A. M. H. Pardede, and S. Syahputra, "ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 1, 2022.
- [32] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 639–648, Mar. 2022, doi: 10.47065/bits.v3i4.1408.
- [33] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *Jurnal Ilmiah MATRIK*, vol. 22, no. 1, 2020.
- [34] B. B. Suherman, "SISTEM PAKAR DIAGNOSA PENYAKIT DAN HAMA PADA TANAMAN JAGUNG MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 2, no. 3, pp. 390–398, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [35] M. Ridho Handoko, "SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [36] D. Alita, I. Sari, and A. Rahman Isnain, "PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK PENDUKUNG KEPUTUSAN PENERIMA BEASISWA," *JDMSI*, vol. 2, no. 1, p. 702022, 2021.
- [37] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.

- [38] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," 2020.
- [39] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, "Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotherapy," vol. 7, pp. 46–50, 2021, [Online]. Available: <http://jurnal.unmer.ac.id/index.php/jtmi>
- [40] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.
- [41] S. Bahri and A. Lubis, "METODE KLASIFIKASI DECISION TREE UNTUK MEMPREDIKSI JUARA ENGLISH PREMIER LEAGUE," vol. 2, no. 1, 2020.
- [42] D. Atika, A. Ari Aldino, S. Informasi, J. Pagar Alam No, L. Ratu, and K. Kedaton, "TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TEKANAN MENTAL PADA MEDIA SOSIAL TWITTER," 2022. [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [43] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [44] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, p. 650, Jul. 2020, doi: 10.30865/mib.v4i3.2181.
- [45] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, p. 1580, Oct. 2021, doi: 10.30865/mib.v5i4.3347.
- [46] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, "Unsupervised Learning Methods for Molecular Simulation Data," *Chemical Reviews*, vol. 121, no. 16. American Chemical Society, pp. 9722–9758, Aug. 25, 2021. doi: 10.1021/acs.chemrev.0c01195.
- [47] H. Sibyan, W. Suharso, E. Suharto, M. A. Manuhutu, and A. P. Windarto, "Optimization of Unsupervised Learning in Machine Learning," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1783/1/012034.
- [48] K. Tyagi, C. Rane, R. Sriram, and M. Manry, "Unsupervised learning," in *Artificial Intelligence and Machine Learning for EDGE Computing*, Elsevier, 2022, pp. 33–52. doi: 10.1016/B978-0-12-824054-0.00012-5.
- [49] D. S. Watson, "On the Philosophy of Unsupervised Learning," *Philos Technol*, vol. 36, no. 2, Jun. 2023, doi: 10.1007/s13347-023-00635-6.
- [50] N. Li, M. Shepperd, and Y. Guo, "A Systematic Review of Unsupervised Learning Techniques for Software Defect Prediction," *Inf Softw Technol*, pp. 1–18, Jul. 2020, [Online]. Available: <http://arxiv.org/abs/1907.12027>
- [51] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. in Springer Texts in Statistics, vol. 103. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-7138-7.
- [52] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 27, Jan. 2022, doi: 10.30865/mib.v6i1.3294.
- [53] S. Mutrofin, T. Wicaksono, and A. Murtadho, "Perbandingan Kinerja Algoritma Kmeans dengan Kmeans Median pada Deteksi Kanker Payudara," *Jurnal Informasi dan Teknologi*, vol. 5, pp. 88–91, Feb. 2023.

- [54] M. N. Zain, "Algoritma Artificial Neural Network dalam Klasifikasi Chest X-Rays Pasien COVID-19," *Jurnal Riset Statistika*, pp. 137–144, Dec. 2022, doi: 10.29313/jrs.v2i2.1426.
- [55] M. Mega Santoni, N. Chamidah, and N. Matondang, "Prediction of Hypertension using Decision Tree, Naïve Bayes and Artificial Neural Networks in KNIME Analytics Platform," 2020.
- [56] D. Galih Pradana, M. L. Alghifari, M. Farhan Juna, and S. Dwisiwi Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 2, pp. 55–60, 2022.
- [57] N. Agustina Purwitasari, M. Soleh, J. Raya Puspiptek, and K. Tangerang Selatan, "Implementasi Algoritma Artificial Neural Network Dalam Pembuatan Chatbot Menggunakan Pendekatan Natural Language Processing (Implementation Of Artificial Neural Network Algorithm In Chatbot Development Using Natural Language Processing Approach)," Feb. 2022.
- [58] C. Nurina Prabiantissa and G. Eka Yuliasuti, "Prediksi Pergerakan Ikan Di Pesisir Pulau Madura Menggunakan Metode Gaussian Mixture Model Dan K-Means Clustering," 2021. [Online]. Available: <https://doi.org/10.25047/jtit.v8i2.244>
- [59] A. Syahfaridzah, A. K. Panggabean, and N. A. Ardiningsih, "MENDETEKSI SECARA OTOMATIS OBJEK GERAKAN BERDASARKAN GAUSSIAN MIXTURE MODEL MENGGUNAKAN APLIKASI MATLAB," *Jurnal METHODIKA*, vol. 6, no. 2, 2020.
- [60] F. Novkaniza and dan Rahmat Al Kafi, "PEMODELAN JUMLAH KASUS BARU HARIAN COVID-19 DI INDONESIA MENGGUNAKAN GAUSSIAN MIXTURE MODEL," *J. Ris. & Ap. Mat*, vol. 07, no. 02, pp. 116–127, 2023.
- [61] R. Sharma, "Study of Supervised Learning and Unsupervised Learning," *Int J Res Appl Sci Eng Technol*, vol. 8, no. 6, pp. 588–593, Jun. 2020, doi: 10.22214/ijraset.2020.6095.