

BAB II

LANDASAN TEORI

2.1 Landasan Teori

2.1.1 Kanker Payudara



Gambar 2. 1 Kanker Payudara

Kanker payudara adalah jenis kanker yang berkembang dari sel-sel yang abnormal atau ganas di dalam jaringan payudara. Ini adalah salah satu jenis kanker yang paling umum di kalangan wanita, meskipun jarang terjadi pada pria. Kanker payudara dapat terjadi baik pada payudara wanita maupun pada payudara pria, namun kebanyakan kasus terjadi pada wanita. Kanker payudara dapat berkembang di berbagai bagian payudara, termasuk kelenjar susu, saluran susu, dan jaringan lemak atau ikat. Kanker payudara dapat dikenali melalui pemeriksaan fisik, mamografi, biopsi, dan pemeriksaan histologi.[9][10] Deteksi dini kanker payudara sangat penting karena dapat meningkatkan peluang kesembuhan dan memungkinkan pengobatan yang lebih efektif.

2.1.2 Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer untuk belajar dari data tanpa harus diprogram secara eksplisit.[11] Dengan machine learning, komputer dapat mengidentifikasi pola-pola yang kompleks dalam data dan menggunakan pola-pola ini untuk membuat prediksi atau pengambilan keputusan. Ada dua jenis utama dari machine learning: supervised learning dan unsupervised learning.[12]

- **Supervised Learning:** Dalam supervised learning, algoritma dipelajari dari data yang sudah dilabeli, yang berarti bahwa data tersebut sudah memiliki pasangan input-output yang ditentukan sebelumnya. Tujuan dari supervised learning adalah untuk membuat model yang dapat mempelajari hubungan antara input dan output sehingga dapat membuat prediksi yang akurat terhadap data baru yang belum pernah dilihat sebelumnya.
- **Unsupervised Learning:** Dalam unsupervised learning, algoritma diberikan data yang tidak dilabeli dan diminta untuk menemukan pola yang bermanfaat dalam data tersebut. Tujuan dari unsupervised learning adalah untuk menggali wawasan yang berguna dari data, seperti mengidentifikasi kelompok-kelompok atau pola-pola yang tersembunyi.

Machine learning telah diterapkan dalam berbagai bidang, termasuk pengenalan wajah, deteksi penipuan keuangan, pengenalan suara, pengobatan medis, dan masih banyak lagi. Dengan kemampuannya untuk belajar dari data, machine learning telah

menjadi alat yang kuat untuk memecahkan masalah kompleks dalam berbagai bidang.

2.1.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) adalah model matematika yang terinspirasi dari struktur jaringan saraf manusia. ANN terdiri dari lapisan-lapisan neuron yang terhubung, yang masing-masing menerima input, melakukan pengolahan, dan mengirimkan output ke lapisan berikutnya.[13] ANN mampu menangani pola-pola yang kompleks dan non-linear dalam data, sehingga cocok untuk klasifikasi dengan fitur-fitur yang kompleks.

2.1.4 k-Nearest Neighbors (KNN)

k-Nearest Neighbors (KNN) adalah algoritma klasifikasi sederhana yang beroperasi dengan cara mencari k tetangga terdekat suatu titik data untuk menentukan kelasnya.[14] Prediksi kelas suatu titik data didasarkan pada mayoritas kelas dari tetangga-tetangganya. KNN sederhana namun dapat memberikan hasil yang baik dalam kasus-kasus dengan pola yang jelas dan kecil sampel data.

2.1.5 K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu teknik evaluasi yang umum digunakan dalam machine learning untuk mengukur kinerja model secara objektif. Teknik ini membagi dataset menjadi k bagian (fold) yang sama besar, di mana salah satu bagian digunakan sebagai data validasi dan sisanya digunakan sebagai data pelatihan. Proses ini diulang sebanyak k kali, di mana setiap kali salah satu dari k bagian digunakan sebagai data validasi dan sisanya sebagai data pelatihan.[15]

Setelah semua iterasi selesai, dilakukan penghitungan rata-rata dari metrik evaluasi yang diinginkan (misalnya, akurasi, presisi, atau recall) dari semua iterasi untuk mendapatkan estimasi yang lebih stabil terhadap kinerja model. Dengan menggunakan K-Fold Cross Validation, kita dapat memastikan bahwa model kita dinilai secara adil dan objektif, karena semua bagian dataset digunakan baik sebagai data pelatihan maupun data validasi. Hal ini membantu menghindari bias yang mungkin muncul jika hanya menggunakan satu set data pelatihan dan validasi tertentu. Selain itu, dengan melakukan iterasi sebanyak k kali, kita dapat memastikan bahwa evaluasi model tidak dipengaruhi oleh pemilihan acak dari bagian data tertentu. K-Fold Cross Validation sangat berguna ketika kita memiliki dataset yang relatif kecil, karena memungkinkan kita untuk memaksimalkan penggunaan data yang tersedia untuk pelatihan dan evaluasi model. Selain itu, teknik ini juga membantu dalam memperkirakan bagaimana model kita akan berperilaku terhadap data yang belum pernah dilihat sebelumnya, yang memungkinkan kita untuk mengukur generalisasi model dengan lebih baik.

2.2 Penelitian Terkait

Tabel 2. 1 Matriks literatur review dan posisi penelitian

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Hasil Penelitian	Kelemahan	Saran
1	Analisis Bibliometrik Penelitian Keputusan Prediksi Payudara	Suhartono1* ; Totok Chamidy2 ; Syahiduz Zaman3 Vol. 7 No. 2 (2023): September	Tujuan dari penelitian ini adalah untuk melakukan analisis bibliometrik terhadap publikasi ilmiah yang membahas penggunaan metode Decision Tree dalam prediksi kanker payudara. Analisis ini mencakup tren penelitian, literatur yang berpengaruh, penulis, negara, serta struktur literatur mengenai topik ini. Penelitian ini juga bertujuan untuk mengidentifikasi tema utama dalam penelitian kanker payudara serta menyoroti pentingnya pengembangan terusan dalam topik ini.	Hasil analisis artikel yang berpengaruh menemukan peningkatan sitasi dan jumlah penulis yang eksponensial dalam penelitian ini pada periode 2005-2023, dimana China sebagai negara yang dominan dalam penelitian. Pada analisis thematic map menghasilkan tiga topik penelitian yaitu bidang kedokteran, bidang komputer dan bidang bioinformatika. Topik penelitian dalam penggunaan metode Decision Tree untuk prediksi kanker payudara adalah termasuk	Salah satu kelemahan penelitian ini adalah kurangnya analisis mendalam terhadap faktor-faktor eksternal yang dapat memengaruhi hasil bibliometrik, seperti perubahan tren penelitian global dan faktor sosial-politik yang dapat memengaruhi jumlah publikasi dan kutipan (Suhartono et al., 2023) [1].	Sebagai saran, penelitian ini dapat diperluas dengan mempertimbangkan faktor-faktor eksternal yang dapat memengaruhi hasil bibliometrik, seperti perubahan tren penelitian global dan faktor sosial-politik yang dapat memengaruhi jumlah publikasi dan kutipan. Hal ini dapat memberikan pemahaman yang lebih komprehensif terhadap konteks penelitian mengenai penggunaan metode Decision Tree dalam prediksi kanker payudara

2	<p>Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Payudara Menggunakan Algoritma Machine Learning</p>	<p>Rian Oktafiani1), Arief Hermawan2), Donny Avianto3) Jurnal Sains dan Informatika i p-ISSN: 2460-173X Volume 9, Nomor 1, Juni 2023</p>	<p>Penelitian ini bertujuan menemukan komposisi data terbaik untuk hasil klasifikasi kanker payudara yang baik</p>	<p>pada bidang komputer. Hasil penelitian menunjukkan performa akurasi yang berbeda pada ketiga algoritma tergantung pada metode validasi. Skema holdout validation dengan rasio 75%:25% menghasilkan akurasi terbaik untuk SVM, yaitu 98.89%. Algoritma Random Forest mencapai akurasi terbaik pada rasio split data 55%:45%, yaitu 95.85%. Namun, Naïve Bayes memiliki performa akurasi yang lebih baik saat menggunakan k-fold cross validation dengan akurasi 93.85%. Metode holdout dengan rasio 75:25 terbukti menghasilkan akurasi terbaik untuk klasifikasi</p>	<p>Kelemahan penelitian ini adalah tidak memberikan informasi tentang ukuran dataset yang digunakan dalam analisis klasifikasi kanker payudara</p>	<p>Saran untuk penelitian selanjutnya adalah memperluas penggunaan algoritma deep learning dan melakukan penelitian terhadap jenis kanker lainnya untuk meningkatkan hasil klasifikasi</p>
---	---	--	--	---	--	--

3	Prediksi Penyakit Payudara Menggunakan Metode Decision Tree	Mellina, Ayu Dian Fitri. 2023	<p>Tujuan penelitian ini adalah meningkatkan deteksi dini kanker payudara melalui analisis data laboratorium menggunakan metode klasifikasi. Dengan demikian, diharapkan dapat meningkatkan peluang kesembuhan pasien melalui diagnosis yang lebih cepat dan akurat.</p>	<p>data payudara menggunakan SVM. Hasil dari penelitian ini menunjukkan bahwa algoritma C5.0 memperoleh hasil klasifikasi terbaik dibandingkan dengan algoritma iterative dichotomiser-3 (ID3) dengan nilai accuracy sebesar 80%, precision sebesar 84,2%, recall sebesar 80%, dan F1 score sebesar 80%. Nilai accuracy sebesar 80% menyatakan bahwa sistem dapat melakukan klasifikasi dengan baik sehingga algoritma C5.0 dapat diterima dan dapat digunakan untuk memprediksi deteksi penyakit kanker payudara</p>	<p>enelitian ini menggunakan Breast Cancer Coimbra Dataset dari tahun 2018. Keterbatasan dataset ini mungkin mempengaruhi generalisasi hasil, terutama jika dataset tersebut tidak mewakili keragaman yang cukup dari populasi kanker payudara secara keseluruhan.</p>	<p>penelitian ini dapat memberikan kontribusi yang lebih besar dalam meningkatkan deteksi dini kanker payudara dan memperbaiki kualitas prediksi penyakit.</p>
---	---	-------------------------------	--	---	--	--

4	Klasifikasi Payudara Menggunakan Neural Network Dan Random Forest	Jamaludin, I., Abdul Kholiq Fajar ² , Muhammad Zaenal Mutaqin ³ , Muhammad Malik Mutoffar ⁴ , Didik Setiyadi ⁵ Jurnal Manajemen Informatika & Sistem Informasi (MISI) Volume 7, Nomor 1, Januari 2024. ISSN 2614-1701 (media cetak) ISSN 2614-3739 (media online)	<p>1. Tujuan penelitian pertama adalah untuk menguji dan membandingkan efektivitas algoritma Neural Network dan Random Forest dalam klasifikasi kanker payudara berdasarkan data klinis dan histopatologi.</p> <p>2. Tujuan penelitian kedua adalah untuk mengevaluasi tingkat akurasi kedua algoritma tersebut dalam mengklasifikasikan kanker payudara.</p> <p>3. Tujuan penelitian ketiga adalah untuk menyarankan penambahan atribut dan data testing serta melakukan perbandingan dengan metode klasifikasi lainnya untuk penelitian selanjutnya.</p>	<p>Data diolah menggunakan alat dengan akurasi 96,11% untuk algoritma Neural Network dan 98,86% untuk algoritma Random Forest, yang menunjukkan bahwa Neural Networks memberikan kinerja yang sedikit dari pada algoritma Random Forest dalam hal sensitivitas dan spesifisitas.</p>	Salah satu kelemahan dari penelitian ini adalah penggunaan dataset yang berasal dari sumber tertentu, dalam hal ini dari Kaggle, yang mungkin memiliki keterbatasan dalam representasi data kanker payudara secara umum. Penggunaan dataset yang tidak representatif dapat mempengaruhi hasil klasifikasi dan generalisasi model yang dibuat.	Saran untuk mengatasi kelemahan penggunaan dataset yang mungkin tidak representatif dengan melakukan pengumpulan data dari berbagai sumber yang lebih luas dan representatif untuk meningkatkan validitas hasil penelitian
5	Perbandingan Teknik Klasifikasi Neural Support Vector Machine, dan	Derisma I, [*] , Fajri Febrian I	Tujuan dari penelitian ini adalah melakukan diagnosis penyakit kanker payudara	Dalam penelitian ini didapat algoritma klasifikasi terbaik yaitu algoritma	Salah satu kelemahan dari SVM adalah bahwa algoritma ini hanya dapat bekerja dengan	Saran saya adalah untuk mempertimbangkan penggunaan kernel SVM jika data yang

	<p>Naive Bayes dalam Mendeteksi Kanker Payudara</p>	<p>BINA INSANI ICT JOURNAL Vol. 7, No. 1, Juni 2020, 53-62 ISSN: 2355-3421 (Print) ISSN: 2527-9777 (Online)</p>	<p>berbasis komputasi yang menghasilkan bagaimana kondisi seseorang kanker berdasarkan akurasi algoritma.</p>	<p>Kernel SVM dengan tingkat akurasi sebesar 98.9 % dan algoritma terendah yaitu Naive Bayes senilai 96.1 %.</p>	<p>baik pada data yang dapat dipisahkan secara linear. Ketika data tersebut tidak linier, diperlukan penggunaan kernel untuk memetakan ulang data ke ruang dimensi yang lebih tinggi</p>	<p>digunakan tidak dapat dipisahkan secara linear, karena kernel dapat membantu memetakan ulang data ke ruang dimensi yang lebih tinggi untuk meningkatkan performa SVM</p>
<p>6.</p>	<p>Analisa Perbandingan Kinerja Algoritma Klasifikasi Untuk Prediksi Penyakit Kanker Payudara</p>	<p>Rudi Hartono, Yusuf Sumaryana, Alwi Nurfaizi (Jurnal Teknologi Informasi) Vol. 7, No. 1, Juni 2023 P-ISSN 2580-7927 E-ISSN 2615-2738</p>	<p>Tujuan dari penelitian ini adalah untuk menemukan algoritma klasifikasi terbaik dalam memprediksi kanker payudara menggunakan bahasa pemrograman Python.</p>	<p>Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki performa akurasi terbaik dalam memprediksi penyakit kanker payudara. Selain itu, hasil perbandingan antara algoritma Decision Tree dan K-Nearest Neighbors (KNN) menunjukkan bahwa Decision Tree memiliki akurasi yang lebih tinggi daripada KNN dalam klasifikasi data pada studi kasus tersebut.</p>	<p>Salah satu kelemahan dari penelitian ini adalah terbatasnya jumlah data sample yang digunakan, sehingga dapat mempengaruhi generalisasi hasil prediksi. Selain itu, penggunaan algoritma klasifikasi yang hanya terfokus pada satu jenis model dapat mengabaikan kemungkinan performa yang lebih baik dari kombinasi model atau teknik lainnya.</p>	<p>Sebagai saran, penelitian ini dapat diperluas dengan menggunakan dataset yang lebih besar dan beragam untuk meningkatkan generalisasi hasil prediksi. Selain itu, mempertimbangkan penggunaan teknik ensemble learning atau kombinasi model untuk meningkatkan performa prediksi kanker payudara. Selain itu, penting untuk melakukan validasi tambahan dan uji coba dengan metode klasifikasi lainnya seperti Support Vector Machine, Naive Bayes, dan K-Nearest Neighbors untuk</p>

							memperoleh pemahaman yang lebih komprehensif tentang performa algoritma klasifikasi dalam mem
--	--	--	--	--	--	--	---