

Analisis Pengaruh Pola Penggunaan Gadget Terhadap Computer Vision Syndrome Menggunakan Algoritma Machine Learning

Analysis of the Influence of Gadget Usage Patterns on Computer Vision Syndrome Using Machine Learning Algorithms

Hamna Izzatunnisa Ahmad^{*1)}, Syahid Abdullah²⁾⁾, Ahmad Chusyairi³⁾

^{1,2,3)}PJJ Informatika, Universitas Siber Asia

Jl. Harsono RM, Ragunan, Pasar Minggu Jakarta Selatan 12550

Riwayat: Copyright ©2023, JITU, Submitted: 16 Februari 2025; Revised: 28 Februari 2024;

Accepted: 29 Maret 2025; Published: 30 Maret 2025

DOI: <https://doi.org/10.32938/jitu.v5i1.9138>

Abstract - This research aims to analyze the impact of gadget usage on eye health using Decision Tree, Random Forest, and Naive Bayes algorithms. The increasing use of gadgets in society potentially causes eye health disorders, specifically Computer Vision Syndrome (CVS) symptoms that require in-depth investigation. Data was collected through a survey questionnaire about gadget usage habits and respondents' eye conditions. The OSEMN method was used to process and analyze data by applying three classification algorithms. Research findings showed the Random Forest algorithm provided the best performance with 73 % accuracy, followed by Naive Bayes at 65 %, and Decision Tree at 64 %. The study provides insights into the impact of gadget usage on eye health and recommendations for maintaining usage balance to prevent health disruptions.

Keywords - Computer Vision Syndrome; Decision Tree; Naive Bayes; Random Forest

Abstrak - Penelitian ini bertujuan menganalisis pengaruh penggunaan gadget terhadap kesehatan mata menggunakan algoritma Decision Tree, Random Forest, dan Naive Bayes. Peningkatan penggunaan gadget di masyarakat berpotensi menimbulkan gangguan kesehatan mata, yaitu gejala Computer Vision Syndrome (CVS) yang memerlukan kajian mendalam. Data diperoleh melalui survei kuesioner tentang kebiasaan penggunaan gadget dan kondisi mata responden. Metode OSEMN digunakan untuk mengolah dan menganalisis data dengan menerapkan tiga algoritma klasifikasi. Hasil penelitian menunjukkan algoritma Random Forest memberikan performa terbaik dengan akurasi 73 %, diikuti Naive Bayes 65 %, dan Decision Tree 64 %. Penelitian memberikan wawasan dampak penggunaan gadget pada kesehatan mata serta rekomendasi menjaga

keseimbangan penggunaan untuk mencegah gangguan kesehatan.

Kata kunci - Computer Vision Syndrome; Decision Tree; Naive Bayes; Random Forest

I. PENDAHULUAN

Penggunaan gadget seperti *smartphone*, tablet, dan laptop telah menjadi bagian tak terpisahkan dari aktivitas sehari-hari masyarakat modern. Penggunaannya memudahkan berbagai aspek kehidupan, namun penggunaan gadget berlebihan dapat menimbulkan masalah kesehatan mata[1]. Studi terkini menunjukkan bahwa paparan berkelanjutan terhadap layar digital dapat menyebabkan *Computer Vision Syndrome* (CVS), sebuah kondisi yang ditandai dengan kelelahan mata digital dengan gejala kompleks seperti mata kering, ketegangan mata, dan potensi kerusakan retina[2], [3].

Risiko CVS semakin meluas karena penggunaan gadget kini mencakup seluruh rentang usia, dari anak-anak hingga dewasa[4]. Gejala CVS telah dikategorikan menjadi empat kelompok utama: astenopia, gangguan permukaan bola mata, gangguan visual, dan gejala ekstraokular[5]. Berbagai penelitian telah mengidentifikasi faktor risiko CVS, termasuk kebiasaan penggunaan gadget seperti jarak terlalu dekat dengan layar, pencahayaan yang buruk, serta durasi penggunaan yang panjang[6]. Faktor lain seperti penggunaan lensa kontak dan kondisi medis tertentu juga berkontribusi terhadap peningkatan risiko CVS[7].

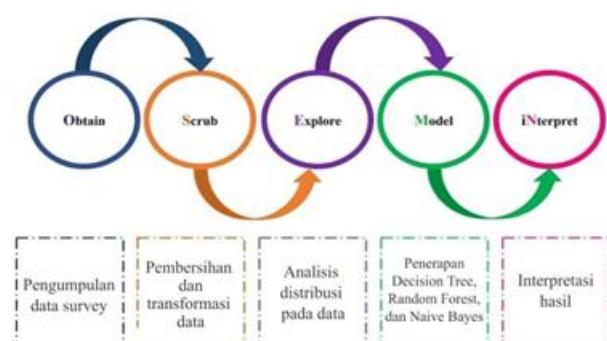
Penelitian sebelumnya mengungkapkan bahwa meskipun banyak studi membahas dampak gadget terhadap kesehatan mata, masih terdapat keterbatasan dalam pendekatan analisis, khususnya penggunaan algoritma berbasis *machine learning* untuk memprediksi risiko CVS. Penelitian-penelitian sebelumnya sebagian besar berfokus pada identifikasi gejala dan faktor risiko, namun belum secara komprehensif menggunakan metode klasifikasi menggunakan *machine learning* untuk memahami pola dan hubungan antara variabel untuk memprediksi risiko CVS.

^{*}) Penulis korespondensi (Hamna Izzatunnisa Ahmad)
Email: 2024.hamnaizzatunnisaahmad@student.unsia.ac.id

Kesenjangan dalam penelitian sebelumnya terletak pada minimnya studi yang mengimplementasikan metode *machine learning* untuk klasifikasi dan prediksi gejala CVS. Oleh karena itu, penelitian ini bertujuan untuk menganalisis faktor-faktor yang berkontribusi terhadap CVS serta mengevaluasi efektivitas algoritma *Decision Tree*, *Random Forest*, dan *Naïve Bayes* dalam mengklasifikasikan gejala CVS berdasarkan data yang diperoleh melalui kuesioner *Computer Vision Syndrome Questionnaire* (CVS-Q). Pemilihan instrumen CVS-Q untuk menilai CVS karena instrumen tersebut telah terbukti valid dan andal dalam mengukur gejala CVS[8]. Dengan menggunakan pendekatan ini, diharapkan penelitian dapat memberikan kontribusi dalam pengembangan model prediksi yang lebih akurat untuk memahami risiko CVS akibat penggunaan gadget secara berlebihan.

II. METODE PENELITIAN

Penelitian ini menggunakan metode *Obtain*, *Scrub*, *Explore*, *Model*, *iNterpret* (OSEMN), yang merupakan kerangka kerja sistematis dalam analisis data[9], [10]. Framework OSEMN mencakup seluruh tahapan analisis data, mulai dari perolehan data, pembersihan, eksplorasi, hingga pemodelan dan interpretasi hasil. Pendekatan ini memastikan pengolahan data yang terstruktur dan terorganisir dengan baik, sehingga menghasilkan model yang andal dan valid[10], [11].



Gambar 1. Framework OSEMN[11]

A. Obtain

Tahap *Obtain* dilakukan dengan mengumpulkan data melalui survei berbasis *Google Form* yang disebarluaskan kepada responden dalam periode tertentu. Metode pengambilan sampel yang digunakan adalah *purposive sampling*, dengan responden yang dipilih berdasarkan kriteria pengguna gadget aktif[12], [13]. Penelitian dilaksanakan pada periode 17-18 Desember 2024, dan mendapatkan 170 data responden yang berhasil dikumpulkan melalui *Google Form*. Kuesioner dalam penelitian ini dirancang secara sistematis dan terdiri dari empat kategori utama dengan total 38 pertanyaan yang terbagi dalam beberapa bagian. Bagian pertama adalah data demografis, mencakup lima pertanyaan yang mengeksplorasi karakteristik responden, meliputi usia (dengan enam kelompok: ≤ 12 , 13-17, 18-24, 25-44, 45-64, ≥ 65 tahun), jenis kelamin, riwayat gangguan mata

seperti rabun atau glaukoma, penggunaan alat bantu penglihatan, dan status aktivitas utama responden. Bagian kedua fokus pada bidang studi atau pekerjaan, dengan dua pertanyaan yang menanyakan bidang spesifik dari 15 kategori yang tersedia serta metode pembelajaran atau bekerja (*offline*, *online*, atau *hybrid*).

Bagian ketiga kuesioner mendetailkan pola penggunaan gadget melalui sembilan pertanyaan komprehensif. Responden diminta memberikan informasi tentang durasi penggunaan harian, jenis gadget utama yang digunakan, waktu penggunaan, tujuan penggunaan, jarak penggunaan layar, posisi saat menggunakan gadget, kondisi pencahayaan lingkungan, keberadaan fitur perlindungan mata, dan tingkat kecerahan layar. Bagian terakhir adalah penilaian gejala CVS, menggunakan instrumen CVS-Q yang mengukur 16 parameter gejala CVS berdasarkan frekuensi dan intensitasnya. Kuesioner ini menilai gejala seperti mata terbakar, mata gatal, sensasi benda asing, mata berair, berkedip berlebihan, mata merah, mata nyeri, kelopak mata berat, mata kering, penglihatan kabur, penglihatan ganda, sensitivitas cahaya meningkat, sulit fokus penglihatan dekat, halo di sekitar objek, penglihatan membekuk, dan sakit kepala[14]. Frekuensi gejala dinilai dengan skala: tidak pernah (0), terkadang (1), dan sering/selalu (2), sedangkan intensitas dibagi menjadi sedang (1) dan berat (2). Skor akhir diperoleh dengan mengalikan nilai frekuensi dan intensitas pada setiap parameter, jika skor total sama dengan 6 atau lebih, mengindikasikan adanya CVS[5], [7].

B. Scrub

Tahap *scrub* merupakan proses pembersihan dan persiapan data mentah agar dapat diproses, meliputi standarisasi data kategoris, *mapping* parameter gejala menggunakan aturan *scoring* CVS-Q, dan konversi nilai numerik. Proses ini bertujuan memastikan data bersih, terstruktur, dan siap untuk analisis lebih lanjut[9], [11], [15].

C. Explore

Data yang telah dibersihkan kemudian dianalisis pada tahap *Explore* menggunakan *Exploratory Data Analysis* (EDA), yang mencakup pemeriksaan properti data, perhitungan statistik deskriptif, serta visualisasi distribusi status CVS, pola gejala, serta pola hubungan antara variabel penggunaan gadget dan gejala CVS[6], [7], [9].

D. Model

Tahap *Model* bertujuan untuk menganalisis data[9]. Pada tahap ini, data yang telah diproses digunakan untuk membangun model klasifikasi dengan algoritma *Decision Tree*, *Random Forest*, dan *Naive Bayes*. Algoritma *Decision Tree* dipilih karena kemampuannya dalam menyederhanakan data kompleks menjadi model yang mudah diinterpretasikan[16], *Random Forest* digunakan untuk mengurangi *overfitting* dan

meningkatkan akurasi dengan metode *ensemble learning*[17], [18], sementara *Naive Bayes* diterapkan karena efisiensinya dalam klasifikasi berbasis probabilitas[18], [19]. Untuk menangani ketidakseimbangan data, *Synthetic Minority Oversampling TEchnique* (SMOTE) digunakan dalam proses pemodelan[20]. Validasi model dilakukan menggunakan metode *K-Fold Cross Validation* untuk mengukur performa model dengan membagi *dataset* menjadi 5 *fold* pelatihan dan validasi secara acak[21].

E. iNterpret

Tahap terakhir, *iNterpret* bertujuan untuk menyajikan hasil analisis data secara jelas dan objektif[9]. Pada tahap ini, evaluasi hasil klasifikasi dilakukan dengan menggunakan *Confusion Matrix* yang membandingkan label prediksi dengan label aktual. *Confusion Matrix* terdiri dari empat komponen: *True Positive* (TP) dan *True Negative* (TN) yang menunjukkan jumlah prediksi benar untuk kasus positif dan negatif, serta *False Positive* (FP) dan *False Negative* (FN) yang menunjukkan jumlah prediksi salah[22]. Evaluasi juga menggunakan *Classification Report*. Evaluasi ini melibatkan metrik *Accuracy* untuk mengukur persentase prediksi yang benar terhadap total data[23], *Precision* untuk menunjukkan rasio prediksi positif yang benar dibandingkan dengan total prediksi positif[11], *Recall* untuk mengukur sejauh mana model mampu mengenali dan mengklasifikasikan dengan benar kasus-kasus yang sebenarnya positif[24], dan *F1-Score* sebagai rata-rata harmonis antara Precision dan Recall untuk dataset tidak seimbang[23]. Evaluasi kinerja model dilakukan untuk menilai kinerja model dalam mengidentifikasi hubungan antara faktor-faktor penyebab CVS dan gejala CVS.

Selain itu, pada tahap *iNterpret* metode *Feature Importance* diterapkan untuk *Decision Tree* dan *Random Forest*, sedangkan *Mutual Information* diterapkan untuk *Naive Bayes* yang bertujuan untuk mengukur kontribusi masing-masing variabel dalam proses klasifikasi[25]. Interpretasi hasil penelitian ini bertujuan untuk memberikan wawasan yang mendalam mengenai dampak pola penggunaan gadget terhadap risiko CVS.

III. HASIL DAN PEMBAHASAN

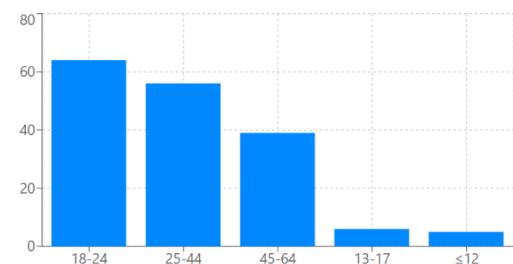
Penelitian ini mengungkapkan beberapa temuan penting tentang pola penggunaan gadget dan dampaknya terhadap CVS yang akan dibahas sesuai urutan *framework OSEMNN*.

A. Obtain

Dari pengumpulan data menggunakan kuesioner, analisis karakter dari 170 responden berdasarkan distribusi hasil pengumpulan data adalah mayoritas responden berasal dari kelompok usia 18-24 tahun dengan jumlah lebih dari 60 responden, yaitu adalah umur-umur mahasiswa atau pekerja muda, kemudian diikuti kelompok usia 25-44 tahun dan 45-64 tahun.

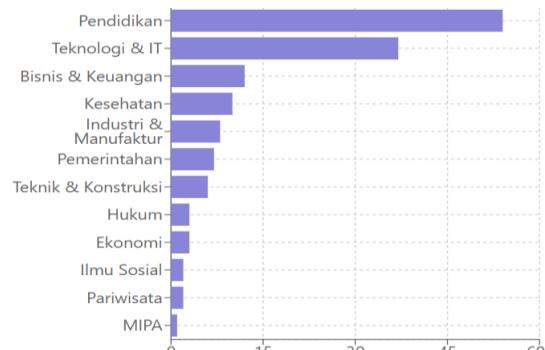
Partisipasi dari kelompok usia 13-17 tahun dan ≤ 12 tahun sangat rendah.

Distribusi Usia Responden



Gambar 2. Distribusi Usia

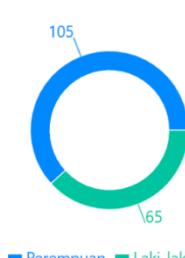
Pendidikan menjadi kategori bidang dengan jumlah responden tertinggi, disusul Teknologi & IT serta Bisnis & Keuangan. Bidang MIPA, Pariwisata, dan Ilmu Sosial memiliki keterlibatan lebih sedikit, menunjukkan survei lebih tersebar kepada responden pada bidang pendidikan dan teknologi.



Gambar 3. Distribusi Bidang

Distribusi jenis kelamin menunjukkan bahwa jumlah perempuan lebih dominan, yaitu sebanyak 105 orang, dibandingkan laki-laki yang hanya 65 orang. Sementara itu, distribusi data riwayat gangguan mata dengan jumlah responden yang memiliki riwayat gangguan mata ("Ya") sebanyak 85 orang, sama dengan jumlah responden yang tidak memiliki riwayat gangguan mata ("Tidak"), yaitu 85 orang. Distribusi ini menunjukkan keseimbangan antara kedua kategori.

Distribusi Jenis Kelamin



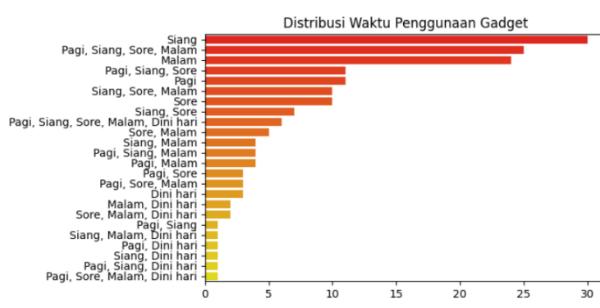
Riwayat Gangguan Mata



Gambar 4. Distribusi Jenis Kelamin dan Riwayat Gangguan Mata

Distribusi waktu penggunaan gadget didominasi oleh waktu "Siang", diikuti kombinasi waktu "Pagi, Siang,

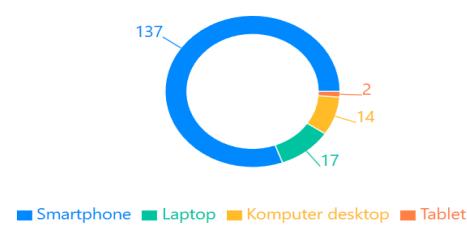
"Sore, Malam" serta waktu "Malam" secara individu. Pola ini mengindikasikan aktivitas gadget lebih intensif saat siang hingga malam, dengan frekuensi yang menurun pada waktu tertentu seperti dini hari. Hal ini mencerminkan kebiasaan pengguna yang lebih aktif menggunakan gadget di waktu yang lebih produktif atau santai, sementara penggunaan pada dini hari lebih jarang terjadi.



Gambar 5. Waktu Penggunaan Gadget

Grafik distribusi jenis gadget yang digunakan dan tujuannya, menunjukkan bahwa *smartphone* menjadi perangkat yang paling dominan, digunakan oleh 137 orang, diikuti oleh laptop (17 orang), komputer desktop (14 orang), dan tablet (2 orang). Hal ini menunjukkan bahwa *smartphone* adalah pilihan utama karena fleksibilitas dan kemampuannya memenuhi berbagai kebutuhan pengguna.

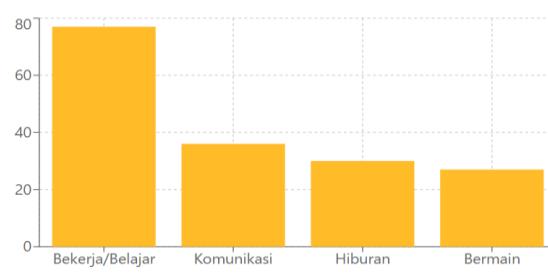
Jenis Gadget yang Digunakan



Gambar 6. Distribusi Jenis Gadget

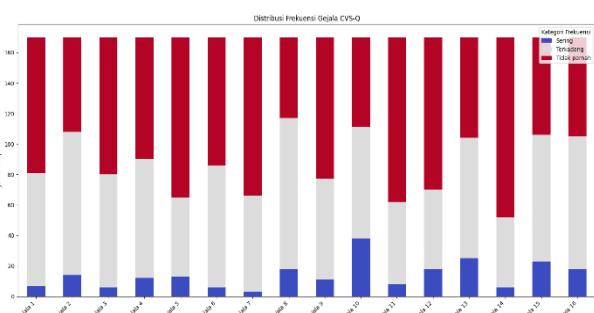
Grafik tujuan penggunaan gadget menunjukkan responden paling sering menggunakan gadget untuk bekerja atau belajar (80 orang), diikuti oleh komunikasi, hiburan, dan bermain dengan jumlah yang lebih kecil. Hal ini menegaskan bahwa kebutuhan produktivitas menjadi prioritas utama, sementara komunikasi dan rekreasi menjadi fungsi tambahan dalam penggunaan gadget.

Tujuan Penggunaan Gadget



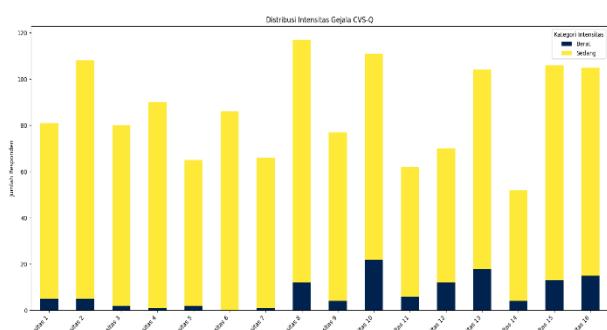
Gambar 7. Distribusi Tujuan Penggunaan Gadget

Berdasarkan visualisasi frekuensi gejala, mata kabur (gejala 10) terlihat paling tinggi (38 responden), sedangkan mata terbakar (gejala 1) paling rendah (7 responden). Pada "Terkadang," kelopak mata berat (gejala 8) paling banyak (99 responden), namun berkedip berlebihan (gejala 5) paling sedikit (52 responden). Kategori jawaban "Tidak Pernah," halo di sekitar objek (gejala 14) paling banyak (118 responden), sedangkan kelopak mata berat (gejala 8) paling sedikit (53 responden). Gejala visual seperti mata kabur dan kelopak mata berat paling umum.



Gambar 8. Distribusi Gejala CVS

Visualisasi intensitas menunjukkan mata kabur paling banyak diisi dengan kategori "Berat" (22 responden), sedangkan pada kategori kelopak mata berat paling tinggi di antara intensitas parameter yang lain adalah kategori "Sedang" (105 responden). Pada parameter mata kemerahan tidak ada responden yang mengisi kategori "Berat". Adapun pada parameter sensasi benda asing adalah paling rendah pada kategori "Sedang" (78 responden) di banding parameter intensitas yang lain. Secara umum, kategori jawaban "Sedang" lebih dominan dari "Berat."



Gambar 9. Distribusi Intensitas CVS

B. Scrub

Hasil mapping menunjukkan bahwa nilai frekuensi dan intensitas semua parameter berhasil dikonversi ke bentuk numerik sesuai skema yang ditetapkan. Kategori "Tidak pernah, Terkadang, dan Sering" pada parameter gejala berhasil dipetakan menjadi nilai numerik 0, 1, dan 2. Sedangkan kategori yang "kosong, Sedang, dan Berat" pada parameter intensitas juga sudah dipetakan menjadi nilai numerik 0, 1, dan 2.

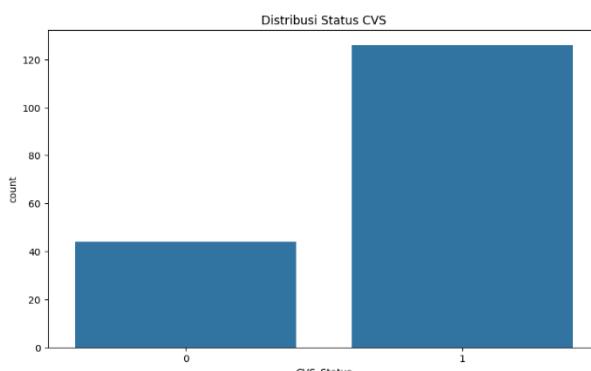
Tabel 1. Mapping Skor CVS

Parameter	Kategori Jawaban	Nilai Numerik
Frekuensi	Tidak Pernah	0
	Terkadang	1
	Sering	2
Intensitas	Kosong	0
	Gejala Sedang	1
	Berat	2

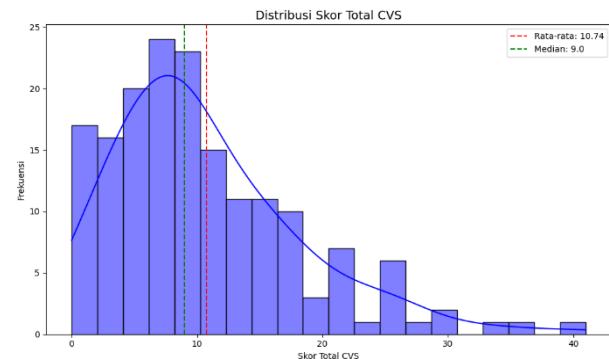
Kolom Skor Total CVS dan CVS_Status, juga berhasil ditambahkan ke *dataset* untuk mendukung analisa lebih lanjut. Kolom CVS status dengan nilai 1 menunjukkan CVS dan nilai 0 menunjukkan tidak CVS.

C. Explore

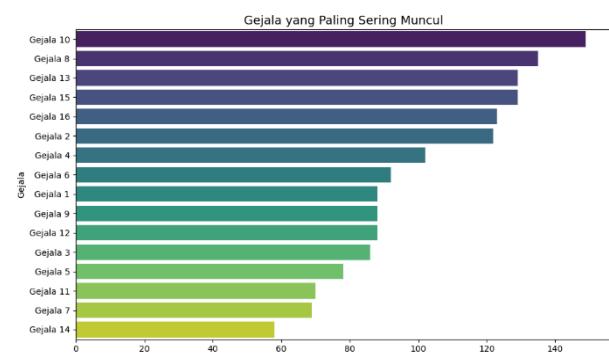
Distribusi status CVS menunjukkan berapa banyak responden yang memiliki gejala CVS dan yang tidak memiliki gejala CVS. Angka 0 menunjukkan tidak CVS, dan angka 1 menunjukkan CVS. Visualisasi di atas menunjukkan perbedaan yang sangat jauh antara responden yang CVS dengan yang tidak CVS, dengan jumlah 126 responden CVS, dan hanya 44 responden yang tidak CVS. Data responden yang CVS dengan yang tidak CVS sangat timpang, yang berarti data tidak seimbang. Sebelum melakukan permodelan, akan dilakukan penyeimbangan data agar tidak mempengaruhi hasil permodelan mengklasifikasikan kategori.

**Gambar 10. Distribusi Status CVS**

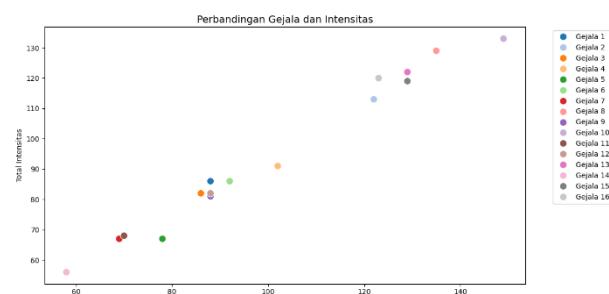
Distribusi skor total CVS dieksplorasi untuk mengetahui total rata-rata skor responden yang menyebabkan lebih banyak data responden CVS daripada yang tidak CVS. Selain itu, juga melihat gejala apa yang paling banyak dialami oleh responden. Distribusi skor total CVS menunjukkan bahwa rata-rata skor berada di angka 10.74 dengan median 9.0, modus 9, dan standar deviasi 7.55, padahal skor ≥ 6 sudah termasuk CVS. Sebagian besar responden memiliki skor di sekitar nilai median dan modus, menunjukkan distribusi yang cenderung terpusat pada rentang menengah. Namun, adanya standar deviasi yang cukup tinggi mengindikasikan bahwa skor bervariasi secara signifikan, dengan beberapa responden memiliki skor yang jauh lebih tinggi atau lebih rendah dari rata-rata.

**Gambar 11. Distribusi Skor Total CVS**

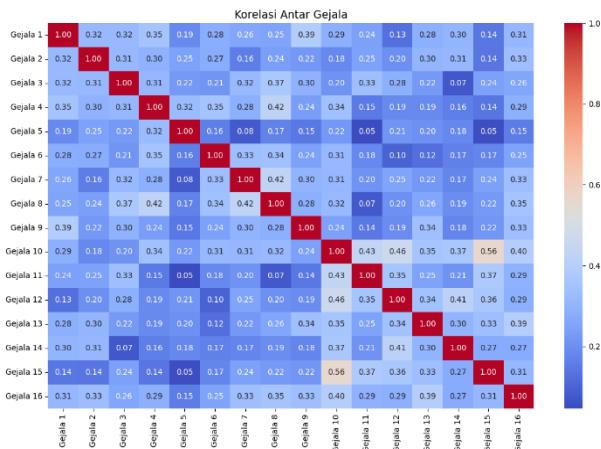
Gejala yang paling sering muncul adalah penglihatan kabur (Gejala 10) dengan frekuensi 149 kali, diikuti oleh kelopak mata terasa berat (Gejala 8) sebanyak 135 kali, sulit fokus pada objek dekat (Gejala 13) dan persepsi penglihatan memburuk (Gejala 15), masing-masing 129 kali. Sebaliknya, gejala yang paling jarang adalah halo di sekitar objek (Gejala 14) dengan hanya 58 kali. Pola ini menunjukkan beberapa gejala lebih sering dirasakan responden dan bisa menjadi indikasi awal CVS.

**Gambar 12. Gejala yang Paling Sering Muncul**

Hasil analisis menunjukkan bahwa frekuensi gejala tidak selalu sebanding dengan intensitasnya. Misalnya, penglihatan kabur (Gejala 10) memiliki frekuensi tertinggi, 149, dengan intensitas 133, sedangkan persepsi penglihatan memburuk (Gejala 15) memiliki frekuensi 129 dengan intensitas 119. Ini menunjukkan bahwa sering munculnya gejala tidak selalu berarti intensitasnya tinggi. Namun, beberapa gejala seperti mata nyeri (Gejala 7) dan mata terbakar (Gejala 1) memiliki frekuensi serta intensitas yang hampir seimbang, menunjukkan hubungan lebih proporsional.

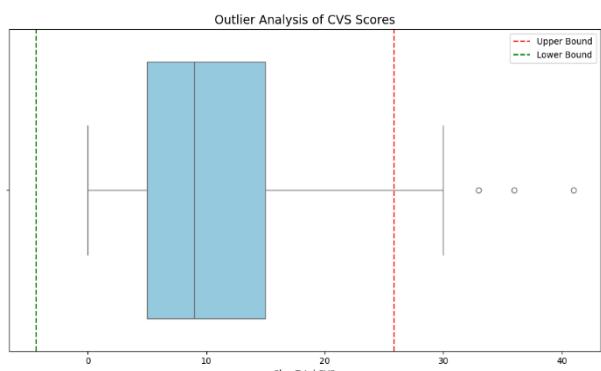
**Gambar 13. Perbandingan Gejala dan Intensitas**

Korelasi antar gejala menunjukkan hubungan positif yang bervariasi. Beberapa pasangan gejala memiliki korelasi kuat, seperti penglihatan kabur (Gejala 10) dan persepsi penglihatan membburuk (Gejala 15) dengan $r = 0.56$, yang menunjukkan keduanya sering muncul bersamaan. Sebaliknya, korelasi rendah atau mendekati nol, seperti berkedip berlebihan (Gejala 5) dan penglihatan ganda (Gejala 11) dengan $r = 0.05$, menunjukkan bahwa keduanya cenderung muncul secara independen atau tidak memiliki hubungan signifikan.



Gambar 14. Korelasi Antar Gejala

Hasil analisis *outlier* pada skor CVS menunjukkan beberapa data memiliki skor yang jauh lebih tinggi dari rata-rata, seperti responden dengan skor 41, 36, dan 33. Semua data ini memiliki status CVS positif (1), mengindikasikan bahwa *outlier* yang ada cenderung berkaitan dengan kondisi CVS positif.



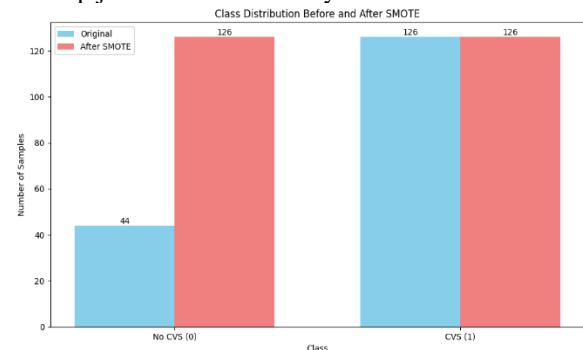
Gambar 15. Outlier Skor CVS

D. Model

Pada tahap ini, model *machine learning* diolah menggunakan tiga algoritma, yaitu *Decision Tree*, *Random Forest*, dan *Naive Bayes*, dengan tujuan membandingkan performa masing-masing model dan menentukan metode yang memberikan hasil paling optimal. *Dataset* dibagi 50:50 menjadi data *training* dan *testing*, karena rasio perbandingan tersebut menghasilkan performa terbaik untuk ketiga algoritma pada kasus ini dibanding rasio perbandingan yang lain.

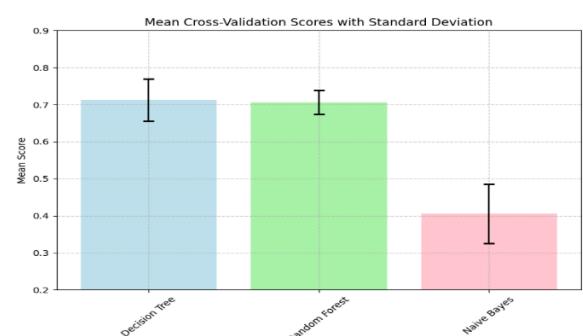
Distribusi kelas pada label CVS_Status menunjukkan ketidakseimbangan dengan 44 sampel kelas 0 dan 126

sampel kelas 1. Ketidakseimbangan ini berpotensi menyebabkan model menjadi bias terhadap kelas mayoritas, sehingga prediksi menjadi kurang akurat untuk kelas minoritas. Setelah penerapan SMOTE, distribusi kelas menjadi seimbang dengan masing-masing kelas memiliki 126 sampel, sehingga model dapat belajar secara lebih adil dan tidak hanya mengutamakan kelas mayoritas dalam proses klasifikasi. SMOTE hanya diterapkan pada data *training*, sementara data *testing* dibiarkan dalam bentuk aslinya agar tetap representatif terhadap jumlah data sebenarnya.



Gambar 16. Perbandingan Distribusi kelas dengan SMOTE

Pada tahap pemodelan, *cross-validation* diterapkan untuk mengevaluasi konsistensi performa model. Hasil validasi silang menunjukkan bahwa *Decision Tree* memiliki performa terbaik dengan skor rata-rata 0.712, *Random Forest* memiliki skor rata-rata 0.706 dengan variabilitas lebih kecil. Sedangkan *Naive Bayes* menunjukkan hasil paling rendah dengan skor rata-rata 0.406, menandakan bahwa metode ini kurang sesuai untuk *dataset* yang digunakan. Hasil ini mengindikasikan bahwa model *Decision Tree* dan *Random Forest* lebih dapat diandalkan dalam mengklasifikasikan status CVS berdasarkan data yang telah diproses dengan SMOTE.



Gambar 17. Cross-Validation 3 Algoritma

Setiap model dilatih menggunakan data *training* yang telah melalui proses SMOTE untuk mengatasi ketidakseimbangan kelas. Setelah pelatihan selesai, model dievaluasi menggunakan data *testing* tanpa proses SMOTE sehingga evaluasi dilakukan pada data dengan distribusi asli. Selain itu, untuk mengukur kinerja model secara keseluruhan, dilakukan proses *cross-validation* pada *dataset* lengkap sebelum model dilatih, guna memastikan kestabilan dan generalisasi model terhadap berbagai *subset* data.

E. Interpret

Tabel berikut menunjukkan perbandingan hasil *confusion matrix* dari ketiga model:

Tabel 2. Confusion Matrix

	Decision Tree	Random Forest	Naive Bayes
<i>True Negative</i> (TN)	4	4	9
<i>False Positive</i> (FP)	16	16	11
<i>False Negative</i> (FN)	15	7	19
<i>True Positive</i> (TP)	50	58	46

Dari tabel di atas, *Random Forest* menunjukkan performa terbaik dengan FN yang paling sedikit (7 kasus) dan TP tertinggi (58 kasus), yang berarti model ini paling mampu mengenali kasus positif CVS dengan baik dibandingkan *Decision Tree* dan *Naive Bayes*. Sedangkan hasil *classification report* untuk ketiga model dapat dirangkum dalam tabel berikut:

Tabel 3. Classification Report

Model-Accuracy	Class	Metric	Score
<i>Decision Tree</i> (64 %)	Non CVS(0)	<i>Precision</i>	0.21
		<i>Recall</i>	0.20
		<i>F1-Score</i>	0.21
	CVS(1)	<i>Precision</i>	0.76
		<i>Recall</i>	0.77
		<i>F1-Score</i>	0.76
<i>Random Forest</i> (73 %)	Non CVS(0)	<i>Precision</i>	0.36
		<i>Recall</i>	0.20
		<i>F1-Score</i>	0.26
	CVS(1)	<i>Precision</i>	0.78
		<i>Recall</i>	0.89
		<i>F1-Score</i>	0.83
<i>Naive Bayes</i> (65 %)	Non CVS(0)	<i>Precision</i>	0.32
		<i>Recall</i>	0.45
		<i>F1-Score</i>	0.38
	CVS(1)	<i>Precision</i>	0.81
		<i>Recall</i>	0.71
		<i>F1-Score</i>	0.75

Analisis perbandingan model menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan akurasi 0.73, diikuti oleh *Naive Bayes* (0.65) dan *Decision Tree* (0.64). *Random Forest* unggul dalam

mendeteksi kasus CVS positif dengan *recall* 0.89, menunjukkan kemampuannya menangkap mayoritas kasus CVS yang ada. Keunggulan ini diperkuat dengan f1-score terbaik (0.83) yang menunjukkan keseimbangan optimal antara *precision* dan *recall*.

Naive Bayes memiliki akurasi keseluruhan yang lebih rendah dari *Random Forest*, tetapi menunjukkan keseimbangan yang lebih baik dalam mendeteksi kasus non-CVS dengan *recall* 0.45, jauh lebih tinggi dibandingkan kedua model lainnya yang hanya mencapai 0.20. Hal ini menunjukkan *Naive Bayes* lebih sensitif dalam mengenali kasus negatif.

Decision Tree, walaupun memiliki akurasi terendah, menunjukkan performa yang konsisten antara *precision* dan *recall* untuk kasus CVS positif (0.76 dan 0.77). Jika mempertimbangkan kompleksitas model dan kecepatan pemrosesan, *Decision Tree* bisa menjadi alternatif yang layak karena memiliki *recall* yang lebih baik dari *Naive Bayes* untuk kasus CVS, meski dengan *accuracy* yang sedikit lebih rendah.

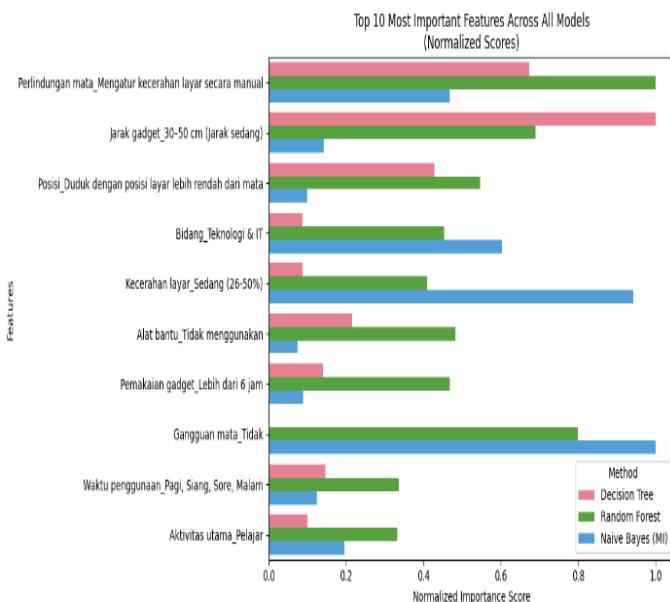
Random Forest merupakan model terbaik untuk untuk menangkap kasus CVS, f1-score terbaik untuk keseimbangan *precision-recall*, dan *accuracy* tertinggi untuk performa keseluruhan. Namun, jika diperlukan keseimbangan yang lebih baik dalam mendeteksi kasus non-CVS, *Naive Bayes* bisa menjadi alternatif yang layak dipertimbangkan.

Setelah melihat evaluasi metrik dari pengolahan data, hal yang perlu diketahui berikutnya adalah fitur-fitur yang mempengaruhi hasil evaluasi metrik data. Terdapat 10 fitur terpenting yang mempengaruhi hasil evaluasi metrik tiga model *machine learning* yang menunjukkan bahwa setiap model memberikan bobot kepentingan yang berbeda untuk masing-masing fitur.

Fitur "Perlindungan mata_Mengatur kecerahan layar secara manual" mendapat perhatian tinggi dari model *Decision Tree* dan *Random Forest* dengan skor normalisasi masing-masing 0.67 dan 1.0. "Jarak gadget_30-50 cm" menjadi fitur terpenting menurut *Decision Tree* dengan skor 1.0, sementara memiliki pengaruh sedang pada *Random Forest* (0.69) dan relatif rendah pada *Naive Bayes* (0.14).

Fitur "Gangguan mata_Tidak" mendapat skor tertinggi dari *Naive Bayes* (1.0) dan cukup tinggi pada *Random Forest* (0.79), namun tidak dianggap signifikan oleh *Decision Tree*. "Kecerahan layar_Sedang (26-50%)" juga menunjukkan variasi penilaian yang signifikan, dengan *Naive Bayes* memberikan bobot tinggi (0.94) sementara *Decision Tree* dan *Random Forest* memberikan bobot yang lebih rendah.

Fitur-fitur lain seperti "Posisi_Duduk dengan posisi layar lebih rendah dari mata", "Bidang_Teknologi & IT", dan "Pemakaian gadget_Lebih dari 6 jam" menunjukkan tingkat kepentingan menengah hingga rendah di berbagai model, mengindikasikan bahwa faktor-faktor ini memiliki pengaruh moderat terhadap CVS. Variasi skor kepentingan antar model menunjukkan bahwa setiap algoritma memiliki cara berbeda dalam mengevaluasi pentingnya suatu fitur, yang mempengaruhi hasil evaluasi metrik.



Gambar 18. Feature Importance

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penggunaan gadget terbukti memiliki pengaruh signifikan terhadap munculnya gejala CVS. Hal ini ditunjukkan oleh 74% responden yang mengalami gejala CVS dengan skor rata-rata 10.74. Gejala yang paling sering dialami meliputi penglihatan kabur, kelopak mata terasa berat, dan kesulitan fokus pada objek dekat. Mayoritas responden menggunakan *smartphone* dengan durasi pemakaian lebih dari 6 jam per hari, terutama untuk keperluan bekerja atau belajar.

Dalam analisis faktor risiko, pengaturan kecerahan layar secara manual dan jarak penggunaan gadget 30-50 cm menjadi faktor yang paling berpengaruh pada *machine learning*. Selain itu, tidak adanya riwayat gangguan mata sebelumnya dan penggunaan kecerahan layar sedang (26-50%) juga merupakan fitur yang berperan pada *machine learning*. Faktor demografis dan perilaku penggunaan gadget seperti posisi duduk, bidang pekerjaan, dan durasi pemakaian menunjukkan tingkat kepentingan menengah hingga rendah pada ketiga model *machine learning*.

Evaluasi performa model menunjukkan bahwa *Random Forest* memiliki performa terbaik dengan akurasi 73%, recall 89%, dan F1-score 0.83 untuk kasus CVS positif. Sementara itu, *Decision Tree* mencapai akurasi 64% dengan keseimbangan antara *precision* dan *recall*, sedangkan *Naive Bayes* memiliki akurasi 65% dan lebih unggul dalam mendeteksi kasus non-CVS. Selain itu, penerapan SMOTE terbukti efektif dalam mengatasi ketidakseimbangan data dan meningkatkan performa model.

Sebagai pengembangan lebih lanjut, penelitian ini dapat ditingkatkan dengan memperbesar ukuran sampel, mengeksplorasi algoritma *machine learning* lain, serta menambahkan variabel penelitian seperti kondisi lingkungan dan pola istirahat guna memperoleh hasil yang lebih komprehensif ini.

DAFTAR PUSTAKA

- [1] A. T. Alisyahbana, “Dampak Penggunaan Gadget Terhadap Kesehatan Mata Remaja,” *Sehat Rakyat: Jurnal Kesehatan Masyarakat*, vol. 2, no. 4, pp. 582–589, 2023, doi: 10.54259/sehatrakyat.v2i4.2272.
- [2] S. Poudel and S. P. Khanal, “Magnitude and Determinants of Computer Vision Syndrome (CVS) among IT Workers in Kathmandu, Nepal,” *Nepalese Journal of Ophthalmology*, vol. 12, no. 24, pp. 245–251, 2020, doi: https://doi.org/10.3126/nepjoph.v12i2.29387.
- [3] A. J. Munsamy *et al.*, “Evidence on the Effects of Digital Blue Light on the Eye: A Scoping Review,” *African Vision and Eye Health*, vol. 81, no. 1, pp. 1–9, 2022, doi: 10.4102/aveh.v81i1.685.
- [4] R. Wandini, L. Novikasari, and M. Kurnia, “Hubungan Penggunaan Gadget Terhadap Kesehatan Mata Anak Di Sekolah Dasar Al Azhar I Bandar Lampung,” *Malahayati Nursing Journal*, vol. 2, no. 4, pp. 810–819, 2020, doi: https://doi.org/10.33024/mnj.v2i4.3049.
- [5] A. W. Azzahro, B. Murti, and E. P. Pamungkasari, “Effect of Digital Device on Computer Vision Syndrome: Meta-Analysis,” *Journal of Epidemiology and Public Health*, vol. 8, no. 2, pp. 181–190, 2023, doi: 10.26911/jepublichealth.2023.08.02.03.
- [6] M. Iqbal, O. Said, O. Ibrahim, and A. Soliman, “Visual Sequelae of Computer Vision Syndrome: A Cross-Sectional Case-Control Study,” *J Ophthalmol*, vol. 2021, pp. 1–16, 2021, doi: 10.1155/2021/6630286.
- [7] F. Ciputra and N. M. Dwipayani, “Computer Vision Syndrome: Sebuah Tinjauan Pustaka,” *Medical Journal : Jurnal Berkala Ilmiah Kedokteran*, vol. 5, no. 1, pp. 49–59, 2022.
- [8] N. C. Sancho *et al.*, “Rasch-Validated Italian Scale for Diagnosing Digital Eye Strain: The Computer Vision Syndrome Questionnaire IT,” *Int J Environ Res Public Health*, vol. 19, no. 8, pp. 1–14, 2022, doi: 10.3390/ijerph19084506.
- [9] F. T. Br Sitepu, V. A. Sirait, and R. Yunis, “Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Prophet Facebook,” *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 1, pp. 99–105, 2021, doi: 10.31294/p.v23i1.9756.
- [10] W. Abbaoui, S. Retal, S. Ziti, B. El Bhiri, and H. Moussif, “Ischemic Stroke Classification Using VGG-16 Convolutional Neural Networks: A Study on Moroccan MRI Scans,” *International journal of online and biomedical engineering*, vol. 20, no. 2, pp. 61–77, 2024, doi: 10.3991/ijoe.v20i02.44845.
- [11] S. Saleem, N. Hasan, A. Khattar, P. R. Jain, T. K. Gupta, and M. Mehrotra, “DeLTran15: A Deep Lightweight Transformer-based Framework for Multiclass Classification of Disaster posts on X,” *IEEE Access*, vol. 12, pp. 153676–153693, 2024, doi: 10.1109/ACCESS.2024.3478790.

- [12] N. Anugrahyanti, M. N. R. Akib, R. N. Maharani, I. D. K. Kamaluddin, S. Latief, and A. I. Arfah, "Pengaruh Penggunaan Laptop dan Gawai terhadap Kesehatan Mata Mahasiswa Kedokteran UMI di Era Pandemi," *Fakumi MedicalJournal: Jurnal Mahasiswa Kedokteran*, vol. 3, no. 10, pp. 750–759, 2023.
- [13] A. Asrulla, R. Risnita, M. S. Jailani, and F. Jeka, "Populasi dan Sampling (Kuantitatif), Serta Pemilihan Informan Kunci (Kualitatif) dalam Pendekatan Praktis," *Jurnal Pendidikan Tambusai*, vol. 7, no. 3, pp. 26320–26332, 2023, [Online]. Available: <https://www.researchgate.net/publication/386875018>
- [14] I. Lapa, S. Ferreira, C. Mateus, N. Rocha, and M. A. Rodrigues, "Real-Time Blink Detection as an Indicator of Computer Vision Syndrome in Real-Life Settings: An Exploratory Study," *Int J Environ Res Public Health*, vol. 20, no. 5, pp. 1–11, 2023, doi: 10.3390/ijerph20054569.
- [15] K. Kumari, M. Bhardwaj, and S. Sharma, "OSEMN Approach for Real Time Data Analysis," *International Journal of Engineering and Management Research*, vol. 10, no. 2, pp. 107–110, 2020, doi: 10.31033/ijemr.10.2.11.
- [16] S. Ghiasi Hafezi *et al.*, "Predicting High Sensitivity C-reactive Protein Levels and Their Associations in A Large Population Using Decision Tree and Linear Regression," *Sci Rep*, vol. 14, no. 30298, pp. 1–16, 2024, doi: 10.1038/s41598-024-81714-2.
- [17] S. Batool and S. Zainab, "A Comparative Performance Assessment of Artificial Intelligence Based Classifiers and Optimized Feature Reduction Technique for Breast Cancer diagnosis," *Comput Biol Med*, vol. 183, pp. 1–19, 2024, doi: 10.1016/j.combiomed.2024.109215.
- [18] T. B. Reza and N. Salma, "Prediction and Feature Selection of Low Birth Weight Using Machine Learning Algorithms," *J Health Popul Nutr*, vol. 43, no. 157, pp. 1–13, 2024, doi: 10.1186/s41043-024-00647-8.
- [19] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Jurnal Informatik*, vol. 18, no. 3, pp. 239–248, 2022.
- [20] P. Patel, S. Chauhan, S. Gupta, T. Gupta, and R. Agrawal, "Mitigating Class Imbalance with Ensemble SMOTEfied-GAN: Advancing Detection Strategies for Automobile Insurance Fraud," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 24, no. 4, pp. 333–342, 2024, doi: 10.5391/IJFIS.2024.24.4.333.
- [21] J. Allgaier and R. Pryss, "Cross-Validation Visualized: A Narrative Guide to Advanced Methods," *Mach Learn Knowl Extr*, vol. 6, no. 2, pp. 1378–1388, Jun. 2024, doi: 10.3390/make6020065.
- [22] M. A. Hossain, T. Hasan, F. Ahmed, S. H. Cheragee, M. H. Kanchan, and M. A. Haque, "Towards Superior Android Ransomware Detection: An Ensemble Machine Learning Perspective," *Cyber Security and Applications*, vol. 3, Dec. 2025, doi: 10.1016/j.csa.2024.100076.
- [23] E. Yeskuatov, S. L. Chua, and L. K. Foo, "Detecting Suicidal Ideations in Online Forums with Textual and Psycholinguistic Features," *Applied Sciences*, vol. 14, no. 21, pp. 1–20, 2024, doi: 10.3390/app14219911.
- [24] Y. Raddad, A. Hasasneh, O. Abdallah, C. Rishmawi, and N. Qutob, "Integrating Statistical Methods and Machine Learning Techniques to Analyze and Classify COVID-19 Symptom Severity," *Big Data and Cognitive Computing*, vol. 8, no. 12, pp. 1–21, 2024, doi: 10.3390/bdcc8120192.
- [25] P. Bathla and R. Kumar, "A Hybrid System to Predict Brain Stroke Using a Combined Feature Selection and Classifier," *Intelligent Medicine*, vol. 4, no. 2, pp. 75–82, 2024, doi: 10.1016/j.imed.2023.06.002.