

PENERAPAN DATA MINING DALAM SELEKSI ATLET SQUASH SUMATERA UTARA DENGAN ALGORITMA C4.5

Muhammad Fachrezzy¹, Insan Taufik², Hermawan Syahputra³, Said Iskandar Al Idrus⁴, Sudianto Manullang⁵, Amansyah⁶

Ilmu Komputer, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

Jl. William Iskandar Ps. V, Medan

E-mail: *muhammadfachrezzy375@gmail.com¹, insan.taufik@gmail.com², hsyahputra@unimed.ac.id³, saidiskandar@unimed.ac.id⁴, amansyah@unimed.ac.id⁵

Abstrak - Olahraga memiliki berbagai tujuan, termasuk rekreasi, prestasi kompetitif, pendidikan, dan latihan jasmani. Squash, sebagai olahraga permainan dengan raket dan bola, diperkenalkan di Indonesia sejak tahun 1960-an. Meskipun belum populer secara luas, squash telah berkembang dengan adanya klub di beberapa provinsi besar. Di Sumatera Utara, squash menjadi salah satu cabang olahraga yang dipertandingkan dalam PON XXI. Dalam proses seleksi atlet squash, penerapan data mining dengan algoritma C4.5 sangat relevan. Algoritma ini digunakan untuk memprediksi pemenang dan membantu dalam menentukan atlet terbaik berdasarkan dataset historis dan atribut yang relevan. Hasil penerapan algoritma menunjukkan keberhasilan dalam memilih atlet yang dapat meningkatkan peluang meraih prestasi. Selain itu, penggunaan algoritma C4.5 memungkinkan pengembangan sistem penseleksian berbasis web yang efisien. Sistem ini memudahkan pengurus provinsi squash Sumatera Utara dalam mengelola seleksi PON secara terstruktur dan transparan. Dengan sistem ini, proses seleksi atlet menjadi lebih cepat dan akurat, sehingga tim pelatih dapat lebih fokus pada pembinaan atlet yang terpilih. Sistem berbasis web juga meningkatkan pengelolaan data dan memungkinkan evaluasi performa atlet secara efektif seiring waktu. Dengan demikian, penerapan teknologi informasi dalam proses seleksi atlet squash tidak hanya meningkatkan efisiensi administratif, tetapi juga mendukung pengembangan potensi atlet dan peningkatan prestasi dalam kompetisi olahraga nasional.

Kata Kunci: Olahraga, Squash, Data mining, Algoritma C4.5, Seleksi atlet, PON XXI.

I. PENDAHULUAN

Olahraga memiliki empat tujuan utama: rekreasi untuk meningkatkan kesehatan fisik dan mental, prestasi kompetitif untuk pencapaian dalam kompetisi, pendidikan olahraga untuk mendukung tujuan pendidikan, dan latihan jasmani untuk kebugaran sehari-hari (Marsheilla Aguss et al., 2022). Olahraga juga mendukung kesehatan dinamis dan statis, namun kehidupan yang sibuk dapat mengurangi aktivitas fisik, meningkatkan stres, dan risiko penyakit kardiovaskular. Aktivitas fisik yang cukup penting untuk mencegah penyakit dan mendukung kesehatan (Badri, 2020).

Squash adalah olahraga yang dimainkan oleh dua orang menggunakan raket dan bola kecil di dalam ruangan yang dibatasi oleh dinding. Tujuan permainan ini adalah memukul bola ke dinding depan sehingga lawan kesulitan mengembalikannya. Ukuran lapangan squash adalah 9,75 meter panjang, 6,40 meter lebar, dengan tinggi dinding depan 4,60 meter dan dinding belakang 2,15 meter (Vai et al., 2022).

Data Mining adalah proses penggalian informasi dan pola berharga dari data besar, melalui tahapan seperti pengumpulan, ekstraksi, analisis, dan statistik. Juga dikenal sebagai *knowledge discovery*, proses ini menghasilkan model atau pengetahuan yang berguna (Arhami dan Nasir, 2020). Kemajuan

dalam data mining didukung oleh perkembangan teknologi informasi, namun tantangan yang muncul adalah fenomena "*rich of data but poor of information*," di mana meskipun banyak data tersedia, penggunaannya belum optimal (Amna et al., 2022).

Algoritma C4.5 adalah metode untuk membangun pohon keputusan, memecah proses pengambilan keputusan yang kompleks menjadi langkah sederhana. C4.5 merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh J. Ross Quinlan, dengan kemampuan menangani nilai hilang, data kontinu, dan *pruning* (Azwanti, 2018). Algoritma ini mengubah dataset besar menjadi struktur pohon keputusan yang dapat dipahami dalam bahasa alami atau diubah menjadi format SQL untuk mencari kategori tertentu (Eska, 2016).

II. TINJAUAN PUSTAKA

Data mining dan *knowledge discovery in database* (KDD) sering digunakan secara bergantian untuk menggambarkan proses penggalian informasi tersembunyi dalam basis data besar, meskipun keduanya berbeda konsep namun sangat terkait. Data mining adalah salah satu tahap dalam keseluruhan proses KDD (Nasari & Darma, 2015). Proses KDD terdiri dari beberapa tahap:

1. Pemilihan Data: Langkah pertama adalah memilih data dari basis data operasional untuk diolah lebih lanjut dan disimpan dalam berkas terpisah.
2. Pra-Pemrosesan: Sebelum data mining, data dibersihkan dari duplikasi, kesalahan, dan inkonsistensi.
3. Transformasi Koding: Data yang telah dipilih diubah agar sesuai untuk proses data mining, disesuaikan dengan pola informasi yang ingin dicari (Yunita, 2018).

Squash dan tenis lapangan memiliki kesamaan asal dari Inggris, namun berbeda dalam tempat dan teknik permainan. Tenis dimainkan di lapangan outdoor yang luas, sementara squash di lapangan indoor yang lebih kecil. Keuntungan squash adalah tidak terpengaruh cuaca karena dimainkan di dalam ruangan. Permainan squash melibatkan gerakan cepat seperti berhenti, perubahan arah, melompat, dan jongkok (Vai et al., 2022).

Algoritma C4.5 adalah metode untuk membangun pohon keputusan, dikembangkan dari algoritma ID3 oleh J. Ross Quinlan. Algoritma ini dapat menangani data numerik dan diskret dengan membagi data secara rekursif berdasarkan atribut yang dipilih, hingga setiap subdivisi terdiri dari data dari kelas yang sama. Pohon keputusan sangat populer karena dapat memecah proses pengambilan keputusan yang kompleks menjadi langkah-langkah sederhana, memudahkan interpretasi dan eksplorasi data. Keunggulannya termasuk kemampuan menangani data noisy, mempelajari ekspresi disjunctive, dan mengidentifikasi hubungan antara variabel input dan target (Safii, 2019).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Dimana:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

Sedangkan rumus untuk menghitung entropi adalah sebagai berikut:

$$Entropy(s) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dimana :

S : himpunan kasus

n : jumlah partisi S

P_i : Jumlah kasus pada partisi ke- i

Sistem adalah kumpulan elemen yang saling terhubung untuk mencapai tujuan tertentu, dengan ciri-ciri seperti memiliki tujuan, batasan, subsistem, keterbukaan, dan saling bergantung. Informasi adalah hasil dari pengolahan data yang memberikan makna dan berguna dalam pengambilan keputusan (Nopriandi, 2018). Perancangan sistem informasi merupakan proses menciptakan sistem baru untuk

menyelesaikan masalah, melibatkan analisis, definisi kebutuhan, dan pengaturan komponen perangkat lunak serta perangkat keras (Nopriandi, 2018).

Hypertext Markup Language (HTML) adalah bahasa dasar untuk pembuatan *web* yang menggunakan tanda untuk menandai berbagai bagian teks. Meskipun HTML digunakan untuk membuat situs web statis yang sederhana, ia sering digabungkan dengan bahasa pemrograman lain untuk meningkatkan fungsionalitas dan interaktivitas situs (Alakel et al., 2019).

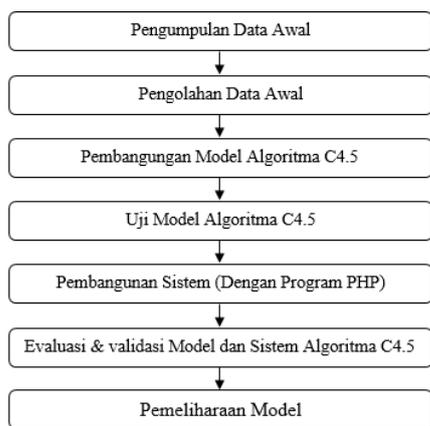
PHP, yang diciptakan oleh Rasmus Lerdorf pada tahun 1995 sebagai *Form Interpreted* (FI), adalah skrip pemrograman untuk membuat situs web (Andy Antonius Setiawan, Arie S.M. Lumenta, 2019). PHP berjalan di sisi server dan mendukung beberapa database, termasuk MySQL dan Oracle. Sebagai perangkat lunak yang dapat diunduh secara gratis, PHP bekerja dengan cara menerima permintaan dari klien melalui kode PHP. Ketika permintaan dikirim, web server mencari berkas yang diminta, dan jika berkas tersebut adalah PHP, server akan memeriksa dan menjalankan skrip tersebut melalui modul PHP. Hasilnya kemudian diubah menjadi kode HTML dan ditampilkan di browser pengguna (Firman et al., 2016).

XAMPP adalah program gratis yang merupakan kompilasi berbagai perangkat lunak, termasuk Apache, MySQL, dan PHP, dan berfungsi sebagai server lokal (*localhost*) yang mendukung berbagai sistem operasi (Andy Antonius Setiawan, Arie S.M. Lumenta, 2019). XAMPP memungkinkan pengembangan halaman web dinamis dengan mudah. Semua programmer yang membuat situs web memerlukan web server untuk menghubungkan file ke basis data, dan beberapa yang umum digunakan termasuk Apache, Wamp, dan XAMPP. Peneliti memilih XAMPP karena memungkinkan instalasi Apache, PHP, dan MySQL secara bersamaan (Saputra & Puspaningrum, 2021).

MySQL, yang diciptakan oleh Michael Monty pada tahun 1979, adalah sistem manajemen basis data relasional (RDBMS) yang digunakan untuk menyimpan berbagai informasi, dari catatan sederhana hingga data besar seperti foto dan video (Peprizal & Syah, 2020). MySQL dikenal sebagai salah satu database paling populer karena kinerjanya yang baik, keandalan, dan kemudahan penggunaan. MySQL memungkinkan penyimpanan, pengelolaan, dan akses data menggunakan SQL, dengan dukungan untuk tabel relasional, indeks, transaksi, dan prosedur penyimpanan. Sebagai perangkat lunak *open-source*, MySQL dapat digunakan secara gratis dan dikembangkan oleh komunitas luas, serta kompatibel dengan berbagai platform, menjadikannya pilihan utama dalam pengembangan aplikasi *web* (Priyandanu et al., 2020).

III. METODE PENELITIAN

Penelitian ini dilakukan di Sekretariat Pengprov PSI Sumatera Utara, di Club House Komplek Cemara Hijau, Jalan Metal Medan, Kecamatan Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara. Metode yang digunakan adalah Waterfall, yang merupakan pendekatan sistematis dan berurutan dimulai dari kebutuhan sistem hingga analisis, desain, koding, pengujian, dan pemeliharaan. Metode Waterfall menawarkan alur hidup perangkat lunak yang terurut, dan penelitian ini akan mengikuti tahapan yang ditunjukkan pada gambar 3 (Wijaya & Astuti, 2019).



Gambar 1. Tahapan Penelitian

Penelitian ini menggunakan data SELEKDA tahun 2023 dengan model Algoritma C4.5 untuk memprediksi hasil seleksi atlet. Tahapan penelitian meliputi:

1. Pengumpulan Data Awal: Mengambil data sekunder dari selekda 2022 yang diperoleh dari Pengprov Squash Sumatera Utara.
2. Praproses Data: Melakukan pembersihan data untuk menghilangkan duplikasi dan kesalahan, serta membagi data menjadi data latih dan uji.
3. Pembangunan Model C4.5: Membangun decision tree berdasarkan data yang telah diproses dengan memilih atribut sebagai simpul, membuat cabang untuk nilai atribut, dan memisahkan kasus hingga kelas menjadi seragam.
4. Uji Model C4.5: Menggunakan model untuk klasifikasi data baru, dengan melakukan pemangkasan untuk menghindari overfitting.
5. Pengembangan Sistem: Mengimplementasikan model prediksi seleksi atlet menggunakan PHP dengan metode waterfall.
6. Evaluasi Model: Mengevaluasi model dengan metrik seperti akurasi dan presisi.
7. Pemeliharaan Model: Memelihara model untuk memastikan kualitas dan performa seiring waktu.

Variabel Data

1. Fisik Pemain: Diukur dari kelentukan, keseimbangan, daya tahan otot, kecepatan, dan lainnya dengan skala 1 (kurang sekali) hingga 5 (baik sekali).
2. Umur: Dianggap ideal antara 18-25 tahun, dengan kategori muda (12-18 tahun) dan dewasa (19-30 tahun).
3. Prestasi: Dinilai dari jumlah medali dengan kategori 0-2 (kurang sekali) hingga 9-10 (tinggi sekali).
4. Kualitas Teknik: Diukur dari jenis pukulan dengan kategori 0-10 (kurang sekali) hingga 41-50 (tinggi sekali).
5. Kedisiplinan: Dinilai dari tingkat kepatuhan terhadap aturan, dengan skala 1 (kurang sekali) hingga 5 (tinggi sekali).

Penelitian ini bertujuan untuk menghasilkan sistem prediksi yang akurat dalam seleksi atlet.

IV. HASIL DAN PEMBAHASAN

Penelitian ini adalah penelitian studi kasus mengenai penerapan algoritma C4.5 untuk seleksi atlet Squash di Team Squash Sumut. Data yang diambil adalah data sekunder berupa data seleksi atlet daerah pada tahun 2023 dan memperoleh 5 parameter yaitu, umur, fisik, Teknik, prestasi dan kedisiplinan. Proses selanjutnya adalah melakukan rekapitulasi terhadap data-data yang telah di kumpulkan. Hasil rekapitulasi tersebut dibuat dalam bentuk tabel 1.

Tabel 1. Rekapitulasi Hasil Seleksi Atlet Squash Sumut

N o	Nama	Asal	U mu r	Fis ik	Pres tasi	Tek nik	Disi plin	Ha sil
1	Eris Setiawan	Deli Serdang	37	4	10	50	3	Ya
2	Wahyu Hidayah	Deli Serdang	22	5	7	40	5	Ya
3	Agung Setiawan	Labuhan Batu Selatan	22	5	5	37	5	Ya
4	Jerry G.K Sinurat	Tapanuli Utara	21	4	5	40	4	Ya
5	Muhammad Syahril	Tanjung Balai	22	4	5	40	4	Ya
6	Muhammad Fachrezzy	Deli Serdang	22	3	6	37	5	Ya
7	Ahmad Djibrin Ansori	Medan	20	3	5	45	4	Ya
8	Fadhilah Aulia Putri	Serdang Berdagai	21	4	7	45	4	Ya

9	Natalia Natasha	Labuhan Batu Utara	21	3	6	50	5	Ya
...

Selanjutnya dari tabel rekapitulasi kolom parameter dilakukan seleksi fitur. Selanjutnya dari tabel rekapitulasi kolom parameter dilakukan seleksi fitur.

Tabel 2. Seleksi Fitur

No	Fisik	Prestasi	Teknik	Disiplin	Hasil
1	tinggi	tinggi sekali	tinggi sekali	cukup	Lolos
2	tinggi sekali	Tinggi	tinggi	tinggi sekali	Lolos
3	sedang	Cukup	tinggi	tinggi sekali	Lolos
4	kurang	kurang sekali	kurang	cukup	Tidak
5	kurang	kurang sekali	kurang	kurang sekali	Tidak
6	sedang	kurang sekali	cukup	kurang	Tidak
7	kurang sekali	kurang sekali	cukup	kurang	Tidak
8	sedang	kurang sekali	kurang	kurang sekali	Tidak
...	sedang	kurang sekali	kurang	cukup	Tidak
67	sedang	kurang sekali	kurang	cukup	Tidak

Kemudian pembangunan model c4.5 mencari entropi dari setiap atribut parameter agar dapat mengevaluasi seberapa baik atribut tersebut dalam memisahkan atau mengelompokkan data menjadi kelas yang berbeda. Entropi digunakan sebagai ukuran ketidakpastian atau kekacauan pada set data, dan nilai entropi yang lebih rendah menunjukkan bahwa atribut tersebut memberikan informasi yang lebih baik dalam membuat keputusan. Dengan demikian, analisis entropi pada setiap atribut bertujuan agar dapat memilih atribut yang paling informatif untuk digunakan dalam pembentukan model atau keputusan lebih lanjut. Berikut adalah perhitungannya dengan contoh parameter fisik:

Rumus perhitungan Entropi:

$$E(S) = -p1 \cdot \log_2(p1) - p2 \cdot \log_2(p2)$$

Keterangan:

$$E(S) = \text{Entropi}$$

$$p1 = \text{Proporsi Lolos}$$

$$p2 = \text{Proporsi Tidak Lolos}$$

Dengan rumus tersebut kita dapat menghitung entropi dari atribut yang ada di setiap parameter pada variabel.

a. Fisik

- **Tinggi Sekali**

Tabel 3. Fisik Parameter Tinggi Sekali

FISIK	Lolos	Tidak Lolos	Total
Tinggi Sekali	2	0	2

Berikut adalah perhitungan parameter Tinggi Sekali:

$$P1 = \frac{\text{jumlah lolos}}{\text{total}}$$

$$P1 = \frac{2}{2}$$

$$P1 = 1$$

Kemudian menghitung P2

$$P2 = \frac{\text{jumlah tidak lolos}}{\text{total}}$$

$$P2 = \frac{0}{2}$$

$$P2 = 0$$

Maka, entropi dari parameter tinggi sekali, yaitu :

$$Entropi (S) = -\frac{1}{67} \cdot \log_2\left(\frac{1}{67}\right) - \frac{0}{67} \cdot \log_2\left(\frac{0}{67}\right)$$

$$Entropi (S) = 0$$

- **Tinggi**

Tabel 4. Fisik Parameter Tinggi

FISIK	Lolos	Tidak Lolos	Total
Tinggi	4	8	12

Berikut adalah perhitungan parameter Tinggi:

$$P1 = \frac{\text{jumlah lolos}}{\text{total}}$$

$$P1 = \frac{4}{12}$$

$$P1 = 0.33334$$

Kemudian menghitung P2

$$P2 = \frac{\text{jumlah tidak lolos}}{\text{total}}$$

$$P2 = \frac{8}{12}$$

$$P2 = 0.6667$$

Maka, entropi dari parameter tinggi, yaitu :

$$E(S) = -\frac{0.33334}{67} \cdot \log_2\left(\frac{0.33334}{67}\right) - \frac{0.6667}{67} \cdot \log_2\left(\frac{0.6667}{67}\right)$$

$$E(S) = 0,104244553$$

- **Sedang**

Tabel 5. Fisik Parameter Sedang

FISIK	Lolos	Tidak Lolos	Total
Sedang	3	24	27

Berikut adalah perhitungan parameter Sedang:

$$P1 = \frac{\text{jumlah lolos}}{\text{total}}$$

$$P1 = \frac{3}{27}$$

$$P1 = 0.1111$$

Kemudian menghitung P2

$$P2 = \frac{\text{jumlah tidak lolos}}{\text{total}}$$

$$P2 = \frac{24}{27}$$

$$P2 = 0.8889$$

Maka, entropi dari parameter sedang, yaitu :

$$E(S) = -\frac{0.1111}{67} \cdot \log_2\left(\frac{0.1111}{67}\right) - \frac{0.8889}{67} \cdot \log_2\left(\frac{0.8889}{67}\right)$$

$$E(S) = 0,09805$$

• Kurang

Tabel 6. Fisik Parameter Kurang

FISIK	Lolos	Tidak Lolos	Total
Kurang	0	23	23

Berikut adalah perhitungan parameter Kurang:

$$P1 = \frac{\text{jumlah lolos}}{\text{total}}$$

$$P1 = \frac{0}{23}$$

$$P1 = 0$$

Kemudian menghitung P2

$$P2 = \frac{\text{jumlah tidak lolos}}{\text{total}}$$

$$P2 = \frac{23}{23}$$

$$P2 = 1$$

Maka, entropi dari parameter kurang, yaitu :

$$E(S) = -\frac{0}{67} \cdot \log_2\left(\frac{0}{67}\right) - \frac{1}{67} \cdot \log_2\left(\frac{1}{67}\right)$$

$$E(S) = 0$$

• Kurang Sekali

Tabel 7. Fisik Parameter Kurang Sekali

Fisik	Lolos	Tidak Lolos	Total
Kurang Sekali	0	3	3

Berikut adalah perhitungan parameter Kurang Sekali:

$$P1 = \frac{\text{jumlah lolos}}{\text{total}}$$

$$P1 = \frac{0}{3}$$

$$P1 = 0$$

Kemudian menghitung P2

$$P2 = \frac{\text{jumlah tidak lolos}}{\text{total}}$$

$$P2 = \frac{3}{3}$$

$$P2 = 1$$

Maka, entropi dari parameter kurang sekali, yaitu :

$$E(S) = -\frac{0}{67} \cdot \log_2\left(\frac{0}{67}\right) - \frac{3}{67} \cdot \log_2\left(\frac{3}{67}\right)$$

$$E(S) = 0$$

Dengan menggunakan hasil entropi yang diperoleh dari setiap atribut fisik, langkah selanjutnya adalah menghitung gain. Proses ini dilakukan untuk mengevaluasi seberapa efektif atribut fisik tersebut dalam mengurangi ketidakpastian atau kekacauan dalam data. Gain diperhitungkan sebagai perbandingan antara entropi awal sebelum pemisahan dengan entropi setelah pemisahan. Tujuan perhitungan gain adalah untuk memilih atribut fisik yang memberikan pemisahan kelas yang optimal, sehingga dapat digunakan dalam pembentukan model atau pengambilan keputusan lebih lanjut.

Dengan rumus:

$$\text{Gain} = E(S) - \sum (I(S || Si) \cdot E(Si))$$

Keterangan:

E(S) = entropi awal (sebelum split) dari dataset S,

|Si| = jumlah sampel pada subset Si setelah split,

|S| = total jumlah sampel sebelum split, dan

E(Si) = entropi setelah split pada subset Si.

Dari rumus tersebut dapat dihitung:

$$\text{Gain} = 0,202294516 - \left(\frac{2}{67} * 0 + \frac{12}{67} * 0,104244553 + \frac{27}{67} * 0,098049963 + \frac{23}{67} * 0 + \frac{3}{67} * 0\right)$$

$$\text{Gain} = 0,144111178$$

Tabel 8. Rekapitulasi Perhitungan Model C4.5 Gain tiap Atribut

PRESTASI	Lolos	Tidak Lolos	Total	Proporsi Lolos	Proporsi Tidak Lolos	Entropi	Gain	Total Kasus
Tinggi Sekali	1	0	1	1	0	0	0,088988	66
Tinggi	2	0	2	1	0	0		
Sedang	6	1	7	0,85714286	0,142857143	0,100546		
Kurang	0	4	4	0	1	0		
Kurang Sekali	0	52	52	0	1	0		
Total	0,100546							
TEKNIK	Lolos	Tidak Lolos	Total	Proporsi Lolos	Proporsi Tidak Lolos	Entropi	Gain	Total Kasus
Tinggi Sekali	2	0	2	1	0	0	0,084093	66
Tinggi	7	7	14	0,5	0,5	0,106733		
Sedang	0	26	26	0	1	0		
Kurang	0	21	21	0	1	0		
Kurang Sekali	0	3	3	0	1	0		
Total	0,106733							

DISIPLIN	Lolos	Tidak Lolos	Total	Proporsi Lolos	Proporsi Tidak Lolos	Entropi	Gain	Total Kasus
Tinggi Sekali	4	1	5	0,8	0,2	0,10252	0,224931	66
Tinggi	4	18	22	0,18181818	0,818181818	0,101946		
Sedang	1	22	23	0,04347826	0,956521739	0,095491		
Kurang	0	11	11	0	1	0		
Kurang Sekali	0	5	5	0	1	0		
Total	0,299957							

$$Gain = 0,954434003 - \left(\frac{2}{37} * 0 + \frac{8}{37} * 0,954434003 + \frac{11}{67} * 0\right)$$

$$Gain = 0,748069894$$

Dapat dilihat pada tabel rekapitulasi variabel disiplin dan prestasi berikut ini:

Tabel 10. Node 1.1 Variabel Disiplin & Prestasi

Node	Disiplin	Jumlah Kasus	Lolos	Tidak Lolos
Prestasi	Tinggi Sekali	0	0	0
	Tinggi	1	1	0
	Sedang	0	0	0

Dari tabel diatas, dapat dipahami bahwa disiplin berperan sebagai node akar dalam pohon Keputusan dengan gain sebesar 0,299957. Selanjutnya, langkah selanjutnya adalah mencari node akar berikutnya.

• **Node 1.1 (Disiplin)**

Jumlah Fisik (Tinggi sekali) dan Disiplin (Tinggi Sekali) = 2
 Jumlah Fisik (Tinggi) dan Disiplin (Tinggi) = 8
 Jumlah Fisik (Sedang) dan Disiplin (Sedang) = 11

Untuk parameter Kurang dan Kurang Sekali tidak dipakai karena memiliki nilai entropi yang rendah di bandingkan dengan parameter lainnya. Dapat kita lihat pada tabel rekapitulasi berikut ini:

Tabel 9. Node 1.1 Variabel Disiplin & Fisik

Node	Disiplin	Jumlah Kasus	Lolos	Tidak Lolos
Fisik	Tinggi Sekali	2	2	0
	Tinggi	8	3	5
	Sedang	11	0	11

Dari tabel 9 dapat di hitung entropi dari tiap tiap parameter seperti berikut:

Berikut adalah perhitungan parameter Tinggi Sekali:

$$E(Tinggi Sekali) = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right)$$

$$E(Tinggi sekali) = 0$$

Berikut adalah perhitungan parameter Tinggi:

$$E(Tinggi Sekali) = -\frac{3}{8} \cdot \log_2\left(\frac{3}{8}\right) - \frac{5}{8} \cdot \log_2\left(\frac{5}{8}\right)$$

$$E(Tinggi sekali) = 0,954434003$$

Berikut adalah perhitungan parameter Sedang:

$$E(Tinggi Sekali) = -\frac{0}{11} \cdot \log_2\left(\frac{0}{11}\right) - \frac{11}{11} \cdot \log_2\left(\frac{11}{11}\right)$$

$$E(Tinggi sekali) = 0$$

Dari entropi tiap parameter berikut ini kita dapat menghitung gain nya:

Dari tabel 10 dapat di hitung entropi dari tiap tiap parameter seperti berikut:

Berikut adalah perhitungan parameter Tinggi Sekali:
 $E(Tinggi Sekali) = -\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right) - \frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$
 $E(Tinggi sekali) = 0$

Berikut adalah perhitungan parameter Tinggi:
 $E(Tinggi Sekali) = -\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right)$
 $E(Tinggi) = 0$

Berikut adalah perhitungan parameter Sedang:
 $E(Tinggi Sekali) = -\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right) - \frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$
 $E(Sedang) = 0$

Dari entropi tiap parameter berikut ini kita dapat menghitung gain nya:

$$Gain = 0 - \left(\frac{0}{1} * 0 + \frac{1}{1} * 0 + \frac{0}{1} * 0\right)$$

$$Gain = 0$$

Hasil $-\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$ dalam perhitungan entropi dianggap 0 karena, meskipun logaritma dari 0 tidak terdefinisi dalam matematika biasa, dalam teori informasi ini diartikan secara berbeda. Jika probabilitas suatu kejadian adalah 0, artinya kejadian tersebut tidak pernah terjadi. Untuk studi kasus kejadian ini adalah kategori Disiplin dan Prestasi yang memiliki nilai Tinggi Sekali. Oleh karena itu, kontribusi informasi atau entropi dari kejadian yang tidak mungkin adalah 0, karena tidak ada ketidakpastian atau informasi yang dihasilkan. Secara matematis, ini dapat dijelaskan dengan konsep limit: $\lim_{x \rightarrow 0} (0 + x \log_2(x)) = 0$. Dengan kata lain, saat probabilitas mendekati 0, hasil ekspresi ini juga mendekati 0, yang sesuai dengan logika bahwa kejadian yang tidak pernah terjadi tidak memberikan informasi tambahan. Jadi, $0 \log_2(0)$ dalam konteks entropi disederhanakan menjadi 0. Dapat dilihat pada tabel 1 rekapitulasi variabel disiplin dan teknik.

Tabel 11. Node 1.1 Variabel Disiplin & Teknik

Node	Disiplin	Jumlah Kasus	Lolos	Tidak Lolos
Teknik	Tinggi Sekali	1	1	0
	Tinggi	6	4	2
	Sedang	9	0	8

Dari tabel 11 dapat di hitung entropi dari tiap tiap parameter seperti berikut:

Berikut adalah perhitungan parameter Tinggi Sekali:

$$E(\text{Tinggi Sekali}) = -\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right)$$

$$E(\text{Tinggi sekali}) = 0$$

Berikut adalah perhitungan parameter Tinggi:

$$E(\text{Tinggi Sekali}) = -\frac{4}{6} \cdot \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \cdot \log_2\left(\frac{2}{6}\right)$$

$$E(\text{Tinggi}) = 0,918295834$$

Berikut adalah perhitungan parameter Sedang:

$$E(\text{Tinggi Sekali}) = -\frac{0}{8} \cdot \log_2\left(\frac{0}{8}\right) - \frac{8}{8} \cdot \log_2\left(\frac{8}{8}\right)$$

$$E(\text{Sedang}) = 0$$

Dari entropi tiap parameter berikut ini kita dapat menghitung gain nya:

$$\text{Gain} = 0,918295834 - \left(\frac{1}{15} * 0 + \frac{6}{15} * 0,918295834 + \frac{8}{15} * 0\right)$$

$$\text{Gain} = 0,769382996$$

Gain tertinggi adalah gain dari Teknik yaitu, 0,769382996 sebagai node selanjutnya.

• **Node 1.2 (Teknik)**

Jumlah Fisik (Tinggi sekali) dan Teknik (Tinggi Sekali) = 0

Jumlah Fisik (Tinggi) dan Teknik (Tinggi) = 4

Jumlah Fisik (Sedang) dan Teknik (Sedang) = 14

Untuk parameter Kurang dan Kurang Sekali tidak dipakai karena memiliki nilai entropi yang rendah di bandingkan dengan parameter lainnya. Dapat kita lihat pada tabel rekapitulasi berikut ini:

Tabel 12. Node 1.2 Variabel Teknik & Fisik

Node	Teknik	Jumlah Kasus	Lolos	Tidak Lolos
Fisik	Tinggi Sekali	0	0	0
	Tinggi	4	3	1
	Sedang	14	0	14

Dari tabel 12 dapat di hitung entropi dari tiap tiap parameter seperti berikut:

Berikut adalah perhitungan parameter Tinggi Sekali:

$$E(\text{Tinggi Sekali}) = -\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right) - \frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$$

$$E(\text{Tinggi sekali}) = 0$$

Hasil $-\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$ dalam perhitungan entropi dianggap 0 karena, meskipun logaritma dari 0 tidak terdefinisi dalam matematika biasa, dalam teori informasi ini diartikan secara berbeda. Jika probabilitas suatu kejadian adalah 0, artinya kejadian tersebut tidak pernah terjadi. Untuk studi kasus kejadian ini adalah kategori Disiplin dan Prestasi yang memiliki nilai Tinggi Sekali. Oleh karena itu, kontribusi informasi atau entropi dari kejadian yang tidak mungkin adalah 0, karena tidak ada ketidakpastian atau informasi yang dihasilkan. Secara matematis, ini dapat dijelaskan dengan konsep limit: $\lim_{x \rightarrow 0} (0 + x \log_2(x)) = 0$. Dengan kata lain, saat probabilitas mendekati 0, hasil ekspresi ini juga mendekati 0, yang sesuai dengan logika bahwa kejadian yang tidak pernah terjadi tidak memberikan informasi tambahan. Jadi, $0 \log_2(0)$ dalam konteks entropi disederhanakan menjadi 0.

Berikut adalah perhitungan parameter Tinggi:

$$E(\text{Tinggi}) = -\frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right)$$

$$E(\text{Tinggi}) = 0,811278124$$

Berikut adalah perhitungan parameter Sedang:

$$E(\text{Sedang}) = -\frac{0}{14} \cdot \log_2\left(\frac{0}{14}\right) - \frac{14}{14} \cdot \log_2\left(\frac{14}{14}\right)$$

$$E(\text{Sedang}) = 0$$

Dari entropi tiap parameter berikut ini kita dapat menghitung gain nya:

$$\text{Gain} = 0,811278124 - \left(\frac{0}{21} * 0 + \frac{4}{21} * 0,954434003 + \frac{14}{21} * 0\right)$$

$$\text{Gain} = 0,656748958$$

Dapat dilihat pada tabel rekapitulasi variabel teknik dan prestasi berikut ini:

Tabel 13. Node 1,2 Variabel Teknik & Prestasi

Node	Teknik	Jumlah Kasus	Lolos	Tidak Lolos
Prestasi	Tinggi Sekali	1	1	0
	Tinggi	2	2	0
	Sedang	0	0	0

Dari tabel 13 dapat di hitung entropi dari tiap tiap parameter seperti berikut:

Berikut adalah perhitungan parameter Tinggi Sekali:

$$E(\text{Tinggi Sekali}) = -\frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right) - \frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right)$$

$$E(\text{Tinggi sekali}) = 0$$

Berikut adalah perhitungan parameter Tinggi:

$$E(\text{Tinggi Sekali}) = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right)$$

$$E(\text{Tinggi}) = 0$$

Berikut adalah perhitungan parameter Sedang:

$$E(\text{Tinggi Sekali}) = -\frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right) - \frac{0}{0} \cdot \log_2\left(\frac{0}{0}\right)$$

$$E(\text{Sedang}) = 0$$

Dari entropi tiap parameter berikut ini kita dapat menghitung gain nya:

$$\text{Gain} = 0 - \left(\frac{1}{3} * 0 + \frac{2}{21} * 0 + \frac{0}{0} * 0\right)$$

$$\text{Gain} = 0$$

Gain tertinggi adalah gain dari Fisik yaitu, 0,656748958 sebagai node terakhir.

Hasil dari perhitungan C4.5 menghasilkan beberapa aturan atau *rule*. Berikut adalah beberapa aturan yang dihasilkan dari analisis pohon keputusan yang telah dilakukan:

Rule 1: Jika "Disiplin" = "Tinggi Sekali," dan "Fisik" = "Tinggi Sekali," maka "Lolos" = "Ya" dengan tingkat kepercayaan 100%.

Rule 2: Jika "Disiplin" = "Tinggi Sekali," dan "Fisik" = "Tinggi," maka "Lolos" = "Ya" dengan tingkat kepercayaan 90%.

Rule 3: Jika "Disiplin" = "Tinggi Sekali," dan "Fisik" = "Sedang," maka "Lolos" = "Ya" dengan tingkat kepercayaan 80%.

Rule 4: Jika "Disiplin" = "Tinggi," dan "Fisik" = "Tinggi," maka "Lolos" = "Ya" dengan tingkat kepercayaan 70%.

Rule 5: Jika "Disiplin" = "Tinggi," dan "Fisik" = "Sedang," maka "Lolos" = "Ya" dengan tingkat kepercayaan 33.33%.

Rule 6: Jika "Disiplin" = "Tinggi," dan "Fisik" = "Kurang," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.

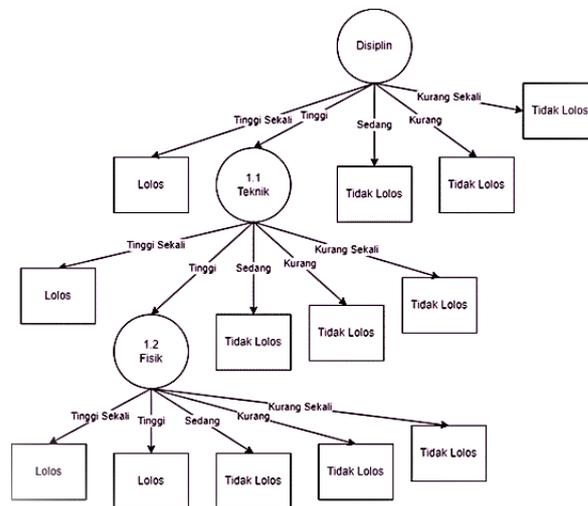
Rule 7: Jika "Disiplin" = "Sedang," dan "Fisik" = "Tinggi Sekali," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.

Rule 8: Jika "Disiplin" = "Sedang," dan "Fisik" = "Tinggi," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.

Rule 9: Jika "Disiplin" = "Sedang," dan "Fisik" = "Sedang," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.

Rule 10: Jika "Disiplin" = "Sedang," dan "Fisik" = "Kurang," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.

Berikut adalah gambar pohon keputusan nya:



Gambar 2. Pohon Keputusan Seleksi At;et

Implementasi gambar keputusan pada data dapat dilihat pada tabel 14.

Tabel 14. Implementasi *Rule*

Nama	Umur	Disiplin	Fisik	Implementasi	Hasil
Wahyu Hidayah	22	tinggi sekali	tinggi sekali	rule 1	lolos
Muhammad Fachrezzy	22	tinggi sekali	tinggi	rule 2	lolos
Melati Anwar	19	sedang	sedang	rule 3	tidak
Muhammad Syahrul	22	tinggi	sedang	rule 4	lolos
Muhammad Izzat	15	sedang	kurang	rule 5	tidak

1. Wahyu Hidayah, Deli Serdang, 22, Disiplin: "tinggi sekali", Fisik: "tinggi sekali"

Implementasi rule :

Rule 1: Jika "Disiplin" = "Tinggi Sekali," dan "Fisik" = "Tinggi Sekali," maka "Lolos" = "Ya" dengan tingkat kepercayaan 100%.

Maka hasil dari implementasi rule ini adalah = Lolos

2. Muhammad Fachrezzy, Deli Serdang, 22, Disiplin: "sedang", Fisik: "tinggi sekali"

Implementasi rule:

Rule 2: Jika "Disiplin" = "Tinggi Sekali," dan "Fisik" = "Tinggi," maka "Lolos" = "Ya" dengan tingkat kepercayaan 90%.

Maka hasil dari Implementasi rule ini adalah = Lolos

- Melati Anwar, Deli Serdang,19, Disiplin:"sedang",Fisik:"sedang" Implementasi rule:

Rule 9: Jika "Disiplin" = "Sedang," dan "Fisik" = "Sedang," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.
Maka hasil dari Implementasi rule ini adalah = Tidak

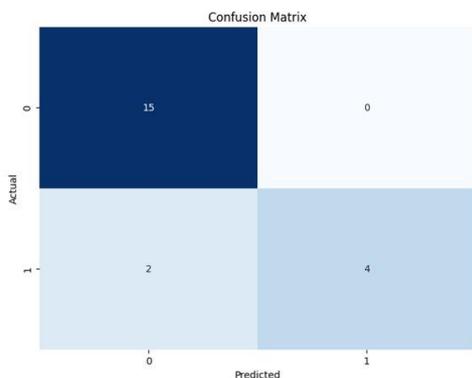
- Muhammad Syahrul, Tanjung Balai,22, Disiplin:"tinggi,Fisik:"tinggi" Implementasi rule:

Rule 4: Jika "Disiplin" = "Tinggi," dan "Fisik" = "Tinggi," maka "Lolos" = "Ya" dengan tingkat kepercayaan 70%.
Maka hasil dari Implementasi rule ini adalah = Lolos

- Muhammad Izzat, Deli Serdang,15, Disiplin:"kurang",Fisik:"sedang" Implementasi rule:

Rule 10: Jika "Disiplin" = "Sedang," dan "Fisik" = "Kurang," maka "Lolos" = "Tidak" dengan tingkat kepercayaan 0%.
Maka hasil dari Implementasi rule ini adalah = Tidak

Setelah melatih model Algoritma C4.5, langkah selanjutnya adalah melakukan prediksi kelas menggunakan data uji. Hasil prediksi ini sangat penting karena memungkinkan untuk mengevaluasi sejauh mana model mampu memprediksi kelas dari data yang belum pernah dilihat sebelumnya. Selanjutnya, untuk mendapatkan wawasan yang lebih mendalam tentang kinerja model, dilakukan perhitungan confusion matrix. Dengan membandingkan nilai prediksi dengan nilai sebenarnya, dapat diidentifikasi True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN). Hasil dari confusion matrix kemudian dicetak ke konsol untuk keperluan analisis lebih lanjut.



Gambar 3. Confusion Matrix dari Model Algoritma C4.5

Pada kasus khusus ini, hasil dari Confusion Matrix menunjukkan bahwa model telah memprediksi 15 kasus sebagai kelas positif dengan benar, dan tidak ada kasus di mana model memprediksi kelas negatif dengan salah. Namun, terdapat 2 kasus di mana model memprediksi kelas negatif ketika seharusnya kelas positif. Analisis ini memberikan wawasan mendalam tentang kinerja model, dan dapat menjadi dasar untuk pengambilan tindakan lebih lanjut dalam mengoptimalkan atau memperbaiki model klasifikasi. Berikut adalah perhitungan metrik dari hasil confusion matrix:

$$accuracy = \frac{TP + TN}{Total} = \frac{15 + 4}{15 + 4 + 0 + 2} = \frac{19}{21} = 0,90 = 90\%$$

$$precision = \frac{TP}{TP + FP} = \frac{15}{15 + 0} = \frac{15}{15} = 1 = 100\%$$

$$recall = \frac{TP}{TP + FN} = \frac{15}{15 + 2} = \frac{15}{17} = 0,88 = 88\%$$

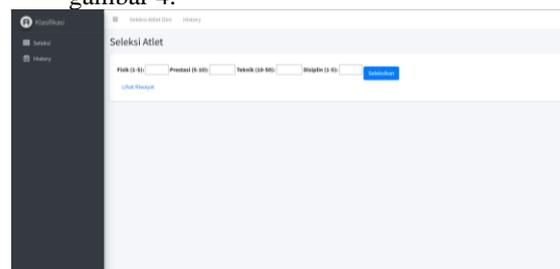
$$True\ Negative\ Rate = \frac{TN}{TN + FP} = \frac{4}{4 + 0} = \frac{4}{4} = 1 = 100\%$$

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} = 2 \times \frac{1 \times 0,88}{1 + 0,88} = 0,93$$

Implementasi Sistem

- Halaman Beranda Website

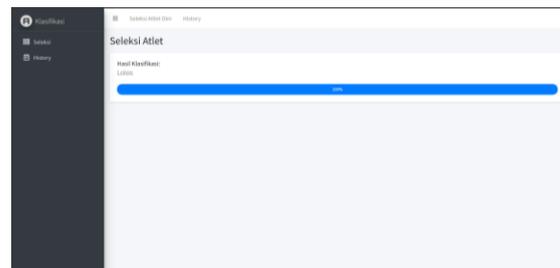
Halaman beranda website ditampilkan pada gambar 4.



Gambar 4. Halaman Beranda Website

- Halaman Hasil Penseleksian Atlet

Halaman hasil Penseleksian website ditampilkan pada gambar berikut.



Gambar 5. Halaman Hasil Penyeleksian Website

- Halaman Histori Parameter & Hasil Penseleksian Website

Gambar 6. Halaman Hasil Histori Parameter & Hasil Penseleksian Website

V. KESIMPULAN DAN SARAN

Kesimpulan

Penerapan Algoritma C4.5 dalam seleksi atlet Squash Sumatera Utara pada ajang SELEKDA PON 2023 berhasil mempermudah proses seleksi dengan mengidentifikasi 9 atlet yang layak dari seluruh kabupaten/kota sebagai atlet Pelatda. Sistem seleksi berbasis web yang dirancang menggunakan algoritma ini memberikan kontribusi positif bagi pengurus provinsi Squash Sumatera Utara dalam meningkatkan efektivitas dan akurasi seleksi atlet untuk ajang tersebut

Saran

Agar sistem seleksi atlet Squash Sumatera Utara berbasis algoritma C4.5 lebih optimal, disarankan untuk menambah fitur atau variabel seleksi serta mengeksplorasi algoritma lain sebagai alternatif. Selain itu, integrasi sistem dengan pengelolaan seleksi atlet secara real-time dapat memberikan pembaruan data yang lebih akurat dan menyediakan informasi yang lebih bermanfaat bagi proses seleksi.

DAFTAR PUSTAKA

Alakel, W., Ahmad, I., & Santoso, E. B. (2019). Sistem Informasi Akuntansi Persediaan Obat Metode First in First Out (Studi Kasus: Rumah Sakit Bhayangkara Polda Lampung). *Jurnal Tekno Kompak*, 13(1), 36. <https://doi.org/10.33365/jtk.v13i1.269>

Amna, S. W., & Sudipa, I. G. I. (2022). *Data Mining*. PT Global Eksekutif Teknologi.

Andy Antonius Setiawan, Arie S.M. Lumenta, S. R. U. A. S. (2019). Rancang Bangun Aplikasi Unsrat E-Catalog. *Jurnal Teknik Informatika*, 14(4), 1-9.

Azwanti, N. (2018). Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu). *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 11-22. <https://doi.org/10.24176/simet.v9i1.1627>

Badri, H. (2020). Aktivitas Olahraga Siswa Sekolah Menengah Atas. *Jurnal Sporta Saintika*, 5(1),

1. <https://doi.org/10.24036/sporta.v5i1.121>

Eska, J. (2016). Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 2(2), 9-13.

Firman, A., F. H., Wowor, & Najoran, X. (2016). Sistem Informasi Perpustakaan Online Berbasis Web. *E-Journal Teknik Elektro Dan Komputer*, 5(2), 29-36.

Marsheilla Aguss, R., Fahrizqi, E. B., Ameraldo, F., Nugroho, R. A., & Mahfud, I. (2022). Perilaku Sportivitas Dan Fairplay Olahraga. *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, 3(2), 204-208. <https://doi.org/10.33365/jsstcs.v3i2.2033>

Nasari, F., & Darma, S. (2015). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama). *Teknologi Informasi Dan Multimedia*, 6-8.

Nopriandi, H. (2018). Perancangan Sistem Informasi Registrasi Mahasiswa. *Jurnal Teknologi Dan Open Source*, 1(1), 73-79. <https://doi.org/10.36378/jtos.v1i1.1>

Peprizal, & Syah, N. (2020). Pengembangan Media Pembelajaran Berbasis Web pada Mata Kuliah Fisika Modern. *Jurnal Ilmiah Pendidikan Dan Pembelajaran*, 4(3), 455-467.

Priyandanu, H., Tabrani, M., Suhardi, & Mutaqin, Z. (2020). Manajemen Persediaan Bahan Baku Berbasis Pada Pt. Tuffindo Nittoku Autoneum Karawang. *Jurnal Ilmiah M-Progress*, 10(1), 90-99. <https://doi.org/10.35968/m-pu.v10i1.370>

Safii, M. (2019). Implementasi Data Mining Dengan Metode Pohon Keputusan Algoritma ID3 Untuk Menentukan Status Mahasiswa. *Jurnal Mantik Penusa*, 2(1), 82-86.

Saputra, A., & Puspaningrum, A. S. (2021). Sistem Informasi Akuntansi Hutang Menggunakan Model Web Engineering (Studi Kasus : Haanhani Gallery). *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(1), 1-7. <http://jim.teknokrat.ac.id/index.php/JTISI>

Vai, A., Adila, F., & Sulastio, A. (2022). Sosialisasi Olahraga Squash Pada Guru Penjasorkes Di Kecamatan Bantan Kabupaten Bengkalis. *IJOsc*, 2(1), 19-28.

Wijaya, Y. D., & Astuti, M. W. (2019). Sistem Informasi Penjualan Tiket Wisata Berbasis Web Menggunakan Metode Waterfall. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, 273-276.

Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru. *Jurnal Sistemasi*, 7(3), 238-249. <https://doi.org/10.32520/stmsi.v7i3.388>