|  | *Jil. 02, No. 01, hlm. 20 – 28 (*  **ISSN 2708-0757** |
| --- | --- |
| JURNALILMU :TREN PENGETAHUANDAN TERAPANTEKNOLOGI )2021  *www.jastt.org* |

Klasifikasi Berdasarkan Algoritma Pohon Keputusan untuk Machine Learning

Bahzad Taha Jijo1\*, Adnan Mohsin Abdulazeez2

*1Jurusan TI, Sekolah Tinggi Teknik Informatika Akre, Universitas Politeknik Duhok, Duhok, Wilayah Kurdistan, Irak, bahzad.taha@dpu.edu.krd*

*2Kepresidenan Universitas Politeknik Duhok, Duhok, Wilayah Kurdistan, Irak, adnan.mohsin@dpu. edu.krd* \**Korespondensi: bahzad.taha@dpu.edu.krd*

***Abstrak***

**pohon keputusan dianggap sebagai metode yang paling terkenal untuk representasi klasifikasi data dari pengklasifikasi. Peneliti yang berbeda dari berbagai bidang dan latar belakang telah mempertimbangkan masalah perluasan pohon keputusan dari data yang tersedia, seperti studi mesin, pengenalan pola, dan statistik. Di berbagai bidang seperti analisis penyakit medis, klasifikasi teks, klasifikasi ponsel cerdas pengguna, gambar, dan banyak lagi, penggunaan pengklasifikasi pohon keputusan telah diusulkan dalam banyak cara. Makalah ini memberikan pendekatan rinci untuk pohon keputusan. Lebih lanjut, spesifikasi makalah, seperti algoritme/pendekatan yang digunakan, kumpulan data, dan hasil yang dicapai, dievaluasi dan diuraikan secara komprehensif. Selain itu, semua pendekatan yang dianalisis**

**dibahas untuk menggambarkan tema penulis dan mengidentifikasi pengklasifikasi yang paling akurat. Akibatnya, penggunaan berbagai jenis kumpulan data dibahas dan temuannya dianalisis.**

***Kata kunci:*** *Machine Learning, Supervised, Klasifikasi, Pohon Keputusan.*

*Diterima: 11 Januari20212021 / Diterima: 15 Maret2021/ Online: 24 MaretI.Dewasa*

PENDAHULUANini

, teknologi sudah banyak berkembang, terutama di bidang Machine Learning (ML), yang berguna untuk mengurangi pekerjaan manusia. Di bidang kecerdasan buatan, ML mengintegrasikan statistik dan ilmu komputer untuk membangun algoritme yang menjadi lebih efisien saat tunduk pada data yang relevan daripada diberi instruksi khusus. Selain pengenalan ucapan, deteksi gambar, pelokalan teks, dll. ML adalah studi tentang algoritma komputasi yang ditingkatkan dari pengalaman secara otomatis. Ini dianggap sebagai subset kecerdasan buatan [1, 2]. Teratur untuk menghasilkan ramalan atau keputusan tanpa secara khusus diprogram untuk melakukannya, algoritma ML membuat populasi model berdasarkan sampel, yang didefinisikan sebagai 'data pelatihan' [3, 4]. Dalam area aplikasi yang luas, seperti pemfilteran email dan visi komputer, algoritma ML digunakan di tempat yang sulit atau tidak praktis untuk membuat algoritma tradisional untuk mengimplementasikan fungsi yang diperlukan [5]. Untuk ML, ada banyak kegunaan, yang paling menonjol adalah data mining prediktif. Dua mekanisme utama dapat dipecah menjadi

pemenuhan klasifikasi ML; pengembangan model dan evaluasi model [6, 7].

Dengan menggunakan kumpulan atribut yang sama, setiap instance di setiap kumpulan data yang digunakan oleh algoritme ML dijelaskan. Atributnya bisa kontinu, kategoris, atau biner [8, 9]. Jika kasus dikenali dengan label yang dikenali (keluaran yang benar), maka pembelajaran disebut terawasi [10, 11]. Pembelajaran Terawasi adalah peran menyimpulkan fungsi dari data pelatihan yang diklasifikasikan adalah pembelajaran mesin. Ini juga menganalisis data pengujian dan membuat tugas turunan yang dapat digunakan sebagai contoh baru untuk dipetakan [12, 13]. Setiap objek input data, bagaimanapun, memiliki label kelas yang telah

ditetapkan sebelumnya. Fungsi utama dari algoritma yang diawasi adalah untuk mempelajari model yang membuat pelabelan yang sama lebih disukai untuk data yang ditawarkan dan mempopulerkan dengan baik pada data yang tidak terlihat. Ini adalah tujuan utama dari algoritma untuk klasifikasi [14].

Klasifikasi mencoba untuk memprediksi kelas tujuan dengan presisi tertinggi. Algoritma klasifikasi mencari hubungan antara atribut input dan atribut output untuk membangun model yang

20 

*doi: 10.38094/jastt20165*

*Tijo & Abdulazeez / Journal of Applied Science and Technology Trends Vol. 02, No. 01, hlm. 20 – 28 (2021)*

proses pelatihan [15 - 17]. Jumlah data yang diperoleh di lingkungan data mining sangat besar [18 - 20]. Jika kumpulan data diklasifikasikan dengan benar dan berisi jumlah node minimum, maka menggunakan metode pohon keputusan optimal [21 - 23].

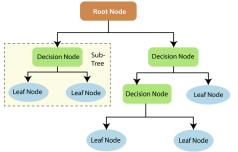
Sebuah pohon keputusan adalah sebuah teknik berbasis pohon di mana setiap jalan mulai dari akar dijelaskan oleh data yang memisahkan urutan sampai hasil Boolean pada node daun tercapai [24 - 27]. Ini adalah contoh hierarki hubungan pengetahuan yang mengandung simpul dan koneksi. Ketika relasi digunakan untuk mengklasifikasikan, node mewakili tujuan [28 - 31].

Dalam makalah ini, tinjauan komprehensif dilakukan untuk pendekatan terbaru dan paling efisien yang telah dilakukan oleh para peneliti dalam tiga tahun terakhir tentang pohon keputusan di berbagai bidang pembelajaran mesin. Juga, rincian metode ini, seperti menggunakan algoritma/pendekatan, kumpulan data, dan temuan yang dicapai diringkas. Selain itu, penelitian ini menyoroti pendekatan yang paling umum digunakan dan metode akurasi tertinggi yang dicapai.

Susunan makalah yang tersisa adalah sebagai berikut: Bagian II berisi algoritme pohon keputusan yang menyebutkan jenis, manfaat, dan kekurangannya; Bagian III memberikan Tinjauan Literatur tentang Algoritma Pohon Keputusan; Bagian IV perbandingan dan pembahasan tentang pohon keputusan, dan bagian terakhir berisi kesimpulan.

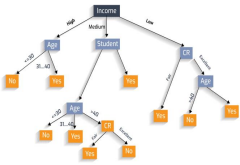
II. POHONKEPUTUSAN LGORITMASalah satuteknik

yang banyak digunakan dalam data mining adalah sistem yang membuat pengklasifikasi [32]. Dalam penambangan data, algoritma klasifikasi mampu menangani sejumlah besar informasi. Ini dapat digunakan untuk membuat asumsi mengenai nama kelas kategorikal, untuk mengklasifikasikan pengetahuan berdasarkan set pelatihan dan label kelas, dan untuk mengklasifikasikan data yang baru diperoleh [33]. Algoritma klasifikasi dalam pembelajaran mesin berisi beberapa algoritma, dan dalam karya ini, makalah difokuskan pada algoritma pohon keputusan secara umum. Gambar 1 mengilustrasikan struktur DT.

Gambar 1. Pohon Keputusan [34]

Pohon keputusan adalah salah satu metode ampuh yang biasa digunakan di berbagai bidang, seperti pembelajaran mesin, pemrosesan gambar, dan identifikasi pola [35]. DT adalah model berurutan yang menyatukan serangkaian tes dasar secara efisien dan kohesif di mana fitur numerik dibandingkan dengan

nilai ambang di setiap tes [36]. Aturan konseptual jauh lebih mudah untuk dibangun daripada bobot numerik dalam jaringan saraf koneksi antara node [37, 38]. Terutama untuk tujuan pengelompokan, DT digunakan. Selain itu, DT adalah model klasifikasi yang biasanya digunakan dalam Data Mining [39]. Node dan cabang terdiri dari setiap pohon. Setiap node mewakili fitur dalam kategori yang akan diklasifikasikan dan setiap subset mendefinisikan nilai yang dapat diambil oleh node [40, 41]. Karena analisisnya yang sederhana dan presisinya pada berbagai bentuk data, pohon keputusan telah menemukan banyak bidang implementasi [42]. Gambar 2 menunjukkan contoh DT.

Gambar 2. Contoh Pohon Keputusan [43]

*A. Jenis-Jenis Algoritma Pohon Keputusan*

Ada beberapa Jenis Algoritma DT seperti: Dikotomi Iteratif 3 (ID3), Penerus ID3 (C4.5), Pohon Klasifikasi Dan Regresi (CART ) [44], CHi-squared Automatic Interaction Detector(CHAID) [45], Multivariate Adaptive Regression Splines (MARS) [46], Generalized, Unbiased, Interaction Detection and Estimation (GUIDE), Conditional Inference Trees (CTREE) [47] ,[48], Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE), Quick, Unbiased and Efficient Statistical Tree (QUEST) [49], [50]. Tabel I menunjukkan perbandingan antara algoritma yang sering digunakan untuk pohon keputusan [51].

*B. Entropi dan Penguatan Informasi*

Entropi digunakan untuk mengukur ketidakmurnian atau keacakan dataset [52], [53]. Nilai entropi selalu terletak antara 0 dan 1. Nilainya lebih baik ketika sama dengan 0 sedangkan lebih buruk ketika sama dengan 0, yaitu semakin dekat nilainya dengan 0 semakin baik. Seperti yang ditunjukkan pada “Gbr. 3”. Jika targetnya adalah �� dengan nilai atribut yang berbeda, maka entropi klasifikasi himpunan �� terhadap �� menyatakan [54], [55]. Seperti yang ditunjukkan pada “persamaan (1)”.

����=1log 2Pi (1)

Entropi( �� ) = P��

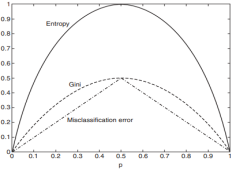
Dimana ����adalah rasio jumlah sampel subset dan nilai ��.

21

*Tijo & Abdulazeez / Jurnal Tren Sains dan Teknologi Terapan Vol. 02, No. 01, hlm. 20 – 28 (2021)*

CHAIDTABEL I:PERBANDINGAN ANTARA ALGORITMA YANG PALING DIGUNAKAN PADA DT

| METODE | CART | C4.5 | Ukuran | QUEST |
| --- | --- | --- | --- | --- |
| yang digunakan untuk pengumpulan variabel input | indeks Gini; Kriteria Twoing | Entropi info-gain | Chi-kuadrat | Chi-kuadrat untuk variabel kategori;kontinu/ordinal |
| - | algoritma | arah | ANOVA | menggunakan |
| variabel | Pemangkasan | Pra | pemangkasan | way |
| satuvariabel | Categorical/ Continuous | Categorical/ Continuous | Categorical/ Continuous | Categorical/ Continuous |
| Split pada setiap node | Biner; Split pada kombinasi linier | Multiple | Multiple | Binary; Membagi kombinasi linier |

TABEL II: BMANFAAT DAN DKEKURANGAN DT 

| Manfaat | Kekurangan |
| --- | --- |
| 1) Mudah dipahami.  2) Cepat diterjemahkan ke satu set prinsip untuk produksi.  3) Dapat mengklasifikasikan hasil kategorikal dan numerik, tetapi atribut yang dihasilkan harus  kategoris.  4) Tidak ada hipotesis apriori yang diambil dengan mempertimbangkan kebaikan hasil. | 1) Mekanisme pengambilan keputusan yang optimal  dapat  dicegah dan  keputusan yang salah dapat mengikuti.  2) Ada banyak lapisan dalam pohon keputusan, yang  membuatnya menarik.  3) Untuk sampel pelatihan yang lebih banyak, kompleksitas perhitungan pohon keputusan  dapat meningkat. |

AKU AKU AKU. LITERATUR RTINJAUAN

Gbr.3. Nilai entropi [56]

Information gain adalah salah satu metrik yang digunakan untuk segmentasi dan sering disebut mutual information. Ini secara intuitif menginformasikan berapa banyak pengetahuan tentang nilai variabel acak [57, 58]. Ini kebalikan dari entropi, semakin tinggi nilainya semakin baik. Gain data ��������( �� , �� ) didefinisikan sebagai berikut pada definisi entropi [59, 60], seperti ditunjukkan pada “persamaan (2)”.

Keuntungan( S , A ) =|Sv|

v V(A) Entropi( Sv ) (2)

|S|

Dimana range atribut �� adalah ��(��), dan ���� adalah himpunan bagian dari himpunan �� sama dengan nilai atribut dari atribut �� [58].

*C. Keuntungan dan Kerugian dari pohon keputusan*

Algoritma DT adalah bagian dari keluarga algoritma pembelajaran terawasi, dan tujuan utamanya adalah untuk membangun model pelatihan yang dapat digunakan untuk memprediksi kelas atau nilai variabel target melalui aturan keputusan pembelajaran yang disimpulkan dari data pelatihan. Algoritma DT dapat digunakan untuk menyelesaikan masalah regresi dan klasifikasi, tetapi memiliki kelebihan dan kekurangan [61 - 63], yang dirangkum dalam Tabel II.

Pohon keputusan digunakan dalam beberapa pembelajaran mesin dan tugas penambangan data sebagai pengklasifikasi. Dalam penelitian ini membahas beberapa karya terbaru tentang DT. Jenis-jenis Tinjauan Pustaka tentang pendekatan DT dirangkum dalam Tabel III.

Zou dkk. [64] Memanfaatkan pohon keputusan (j48), Random Forest (RF), dan algoritma jaringan saraf untuk prediksi diabetes mellitus. Dataset adalah data penelitian fisik untuk rumah sakit di Luzhou, China. Ada 14 karakteristik yang terlibat. Array pelatihan secara acak mengekstraksi data dari 68994 pasien manusia dan diabetes yang stabil, masing-masing. Mereka menggunakan signifikansi penuh minimum Redundancy Maximum Relevance (mRMR) dan Principal Component Analysis (PCA) untuk meminimalkan dimensi. Dalam beberapa hal, efek RF, yang berlawanan satu sama lain, tampaknya lebih tinggi daripada pengklasifikasi lainnya. Juga, 0,8084 adalah hasil terbaik dalam pengumpulan data Luzhou.

Assegie and Nair [65], Memanfaatkan proses klasifikasi DT untuk mengklasifikasikan digit tulisan tangan dari kumpulan data standar digit kaggle dan memperkirakan keakuratan model untuk setiap digit dari 0 hingga 9. Fitur kaggle mencakup 42.000 baris dan 720 kolom yang digunakan untuk pelatihan mesin, fitur vektor digunakan untuk piksel gambar digital. Mereka menggunakan bahasa yang sangat efisien bernama "pemrograman python" untuk penerapan algoritme pembelajaran mesin untuk memetakan grafik tingkat keberhasilan pengklasifikasi dalam realisasi angka tulisan tangan. Temuan

22

*Tijo & Abdulazeez / Jurnal Tren Sains dan Teknologi Terapan Vol. 02, No. 01, pp. 20 – 28 (2021)*

menyatakan bahwa akurasi 83,4% dan pengklasifikasi pohon keputusan berdampak pada pengenalan angka tulisan tangan.

De Felice dkk. [66] menyarankan algoritma pohon keputusan untuk mengenali indikasi klinis yang diketahui dan baru sebelum pengobatan untuk bertahan hidup di Kanker Rektal Lanjutan Lokal (LARC). Analisis menunjukkan bahwa bahkan non-ahli di lapangan, khususnya pohon klasifikasi, dapat dengan mudah menafsirkan proses pembelajaran mesin berbasis pohon. Kesalahan validasi perlu dikelola bahkan untuk mencapai kapasitas statistiknya. Sekitar tahun 2007 dan 2014, pasien dengan LARC

yang dikonfirmasi secara histologis diperiksa datanya. Pendekatan Kaplan-Meier telah digunakan untuk menentukan kelangsungan hidup secara keseluruhan (OS). Ini melibatkan total 100 pasien. 76,4% dan 71,3% adalah poin OS 5 tahun dan 7 tahun. Usia, komorbiditas, ukuran tumor, klasifikasi Tumor Klinis (CT), dan klasifikasi simpul klinis merupakan variabel prediktif penting untuk komposisi pohon (CN). Hasil penelitian menunjukkan bahwa tingkat kelangsungan hidup tertinggi pada pasien usia lanjut dengan ukuran tumor kurang dari 5 cm dan pasien di bawah usia 65 tahun yang memiliki cT3. Pohon keputusan adalah cara untuk mendapatkan pengambilan keputusan praktik klinis yang lebih baik, berdasarkan kumpulan data yang luas.

TABELIII :RINGKASAN PUSTAKATINJAUAN ALGORITMATERKAIT DT Ref

| . | Tahun | Dataset | Teknik | Akurasi |
| --- | --- | --- | --- | --- |
| Nandhini dan KS [75] | 2020 | UCI | DT , KNN , LR , SVM dan NB | DT: 99,93% , KNN: 99,93% , LR: 93,13% , SVM: 90,76% dan NB: 79,52%. |
| Nagra dkk. [79] | 2020 | UCI | SIW-APSO-LS | SIW-APSO-LS: 99,88% |
| Kuang et al. [71] | 2020 | sSCC | Pathan | mengurangi kompleksitas komputasi rata-rata sebesar 47,62% |
| et al. [78] | 2020 | Gambar | Disk Optik (OD)  Segmentasi | OD: 99,61%. |
| Batitis dkk. [73] | 2020 | image | DT | DT: 89,31% |
| Ramadhan dkk. [72] | 2020 | CICIDS2017 | DT dan KNN | DT: 99,91% dan KNN: 98,94% |
| Arowolo et al. [77] | 2020 | RNA-seq Malaria | KNN dan DT | KNN: 86,7% dan DT: 83,3%. |
| De Felice dkk. [65] | 2020 | Pasien dengan LARC yang terbukti secara histologis antara 2007 dan 2014 data mereka | Metode Kaplan-Meier | Tingkat OS 5 tahun: 76,4% Tingkat OS 7 tahun: 71,3%, |
| Zhang et al. [74] | 2019 | Perokok dari Pusat Pengendalian dan Pencegahan Penyakit Tiongkok | DT(XGBoost) dan RF | DT: 84,11% dan RF: 58,11%. |
| Sathiyanarayanan dkk. [82] | 2019 | Dataset Kanker Payudara Wisconsin | DT dan KNN | DT: 99%.  KNN: 97%, |
| Hu dkk. [67] | 2019 | UCI dan COMPAS | OSDT dan BinOCT | UCI / OSDT: 66,90%  COMPAS/ OSDT: 82,881% dan BinOCT: 76,722% |
| Patil dan Kulkarni,  [68] | 2019 | UCI | DST , PT dan MLT | DST: 99,9% terbaik dan terburuk 81,445% |

Sarker dkk. [67] menyajikan Pohon Keputusan Perilaku bernama "BehavDT" struktur konteks-sadar yang memperhitungkan generalisasi berorientasi perilaku konsumen sesuai dengan tingkat pilihan pribadi. Dalam kasus asosiasi yang luar biasa, model BehavDT memberikan keputusan yang komprehensif serta keputusan khusus konteks. Eksperimen dilakukan pada dataset smartphone nyata dari pengguna individu melalui efisiensi model BehavDT. Hasil penelitian menunjukkan bahwa model context-aware Behav DT, yang akurasinya hingga 90%, adalah model yang paling energik dibandingkan dengan model pembelajaran mesin konvensional lainnya.

Hu dkk. [68] menggambarkan algoritma praktis pertama untuk mengoptimalkan pohon keputusan untuk variabel biner. Algoritme adalah co-desain batas analitis yang melibatkan perpustakaan vektor bit khusus dan struktur data yang meminimalkan area pencarian dan teknologi aplikasi saat ini. Mereka menggunakan metode Binary Optimal Classification Trees (BinOCT), yang merupakan metode yang tersedia untuk umum saat ini, untuk menilai akurasi dan membandingkannya dengan Optimal Sparse Decision Trees (OSDT). Mereka juga menggunakan kumpulan data teks dari University of California, Irvine (UCI) Machine Learning Repository dan kumpulan data numerik dari kumpulan data ProPublica COMPAS lainnya. Temuan menunjukkan bahwa ketika dataset COMPAS, pohon keputusan optimal yang dihasilkan oleh OSDT, akurasinya 66,90 %. Selain itu, ketika BinOCT dan OSDT menghasilkan dataset UCI,

pohon keputusan, akurasinya masing-masing adalah 76,722%, 82,881%.

Patil dan Kulkarni [69] memperkenalkan Distributed Spark Tree (DST) untuk mengeksekusi algoritma DT dengan lebih baik dalam hal waktu konstruksi model tanpa kehilangan akurasi. Selain itu, mereka menyarankan untuk menggunakannya dalam iklim Spark. Data dalam arsitektur bersama Spark tidak melakukan eksekusi paralel horizontal. Spark berfungsi dengan baik dan koheren dalam komputasi dalam memori, RDD, dan pengurangan peta. Dataset yang digunakan dari repositori UCI ML dan empat kelas dipilih. File data lebar digunakan untuk menguji kinerja terkait waktu pembuatan model untuk DST, PySpark (PT), dan MLLib (MLT). Temuan menunjukkan bahwa dalam hal akurasi, DST berkinerja lebih baik daripada PT dan MLT, dengan nilai terendah adalah 81,445% dan tertinggi menurut skala dataset adalah 99,9%.

Husain dkk. [70] menawarkan pendekatan modern, yaitu Pohon Keputusan Label Piksel (PLDT), dan memeriksa apakah dapat mencapai efisiensi segmentasi tulang paha yang lebih rinci dalam pencitraan DXA. PLDT meliputi ekstraksi dan seleksi sifat. PLDT digunakan untuk mengungkap pola rahasia yang ditemukan dalam gambar DXA berbeda dengan gambar fotografi. Untuk memutuskan set fitur terbaik untuk model, PLDT menghasilkan tujuh set fitur baru dan menggunakan Global Threshold (GT), Region Growing Threshold (RGT), dan Artificial Neural Networks (ANN). Hasil penelitian mengungkapkan bahwa dalam segmentasi citra DXA, PLDT

23

*Tijo & Abdulazeez / Journal of Applied Science and Technology Trends Vol. 02, No. 01, hlm. 20 – 28 (2021)*

melebihi teknik partisi konvensional lainnya. Untuk masing-masing algoritma seperti PLDT ini akurasinya sebesar 91,4%, GT sebesar 68,4%, RGT sebesar 76%, dan ANN sebesar 84,4%.

Linty dkk. [71] mengusulkan pendekatan baru yang mempengaruhi amplitudo sinyal dari Sistem Satelit Navigasi Global (GNSS) dan digunakan untuk mendeteksi peristiwa kilau ion yang berkaitan dengan akurasi, keandalan, dan kesiapan. Kumpulan luas data pasca-korelasi 50 Hz disediakan oleh penerima GNSS. Hasil penelitian menunjukkan bahwa metode ini, dalam hal akurasi dan F-score, melebihi teknik mutakhir dan dapat mencapai standar yang digerakkan oleh manusia, yaitu tingkat anotasi manual. Ini meningkat pesat karena memperoleh 98% identifikasi, sangat mirip dengan klasifikasi yang digerakkan oleh manusia yang digerakkan dengan tangan.

Kuang dkk. [72] Mengusulkan struktur berdasarkan pohon keputusan bernama Screen Content Coding ( SCC) untuk membuat keputusan cepat dalam situasi dengan menguji fitur yang berbeda dalam set pelatihan. Selain itu, untuk mencegah proses pencarian yang menyeluruh, susunan pohon keputusan yang berurutan diilustrasikan. Selain itu, SCB digunakan sebagai kumpulan data untuk menyeimbangkan SCC dengan mode Intra Block Copy (IBC) dan PaleTte (PLT). Hasilnya menunjukkan bahwa sistem SCC menawarkan penurunan rata-rata 47,62% dalam kompleksitas komputasi, dengan 1,42% kecil di Bjøntegaard delta bitrate (BDBR).

Ramadhan dkk. [73] Mendemonstrasikan analisis perbandingan akurasi dan panjang proses untuk setiap algoritma yang dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dan Decision Tree (DT) untuk mendeteksi serangan DDoS. Selain itu, mereka menggunakan dataset CICIDS2017 yang terdiri dari serangan terbaru dan paket global, standar dan berlaku untuk data dunia nyata dalam format PCAP. Hasil penelitian menunjukkan bahwa akurasi DT untuk mendeteksi serangan DDoS lebih tinggi dari nilai KNN, akurasi DT sebesar 99,91%, dan akurasi KNN sebesar 98,94%.

Batitis dkk. [74] menyajikan sebuah sistem untuk mengidentifikasi hingga 10 sel darah merah yang tidak teratur dan untuk mengetahui tingkat akurasi untuk semua sel darah merah yang abnormal. Selain itu, Untuk mendeteksi sel darah merah yang tidak teratur, mereka menggunakan algoritma DT dalam pemrosesan gambar dan menggunakan bingkai mantan pasien untuk skema di rumah sakit. Selain itu, kamera digunakan untuk memasukkannya ke dalam perangkat lunak untuk menangkap slide. Hasil penelitian menunjukkan bahwa tingkat akurasi rata-rata 89,31% dan tingkat kesalahan rata-rata 10,69%. Selanjutnya, ketidakteraturan sentral dari pucat Codocyte ditemukan menjadi penyebab kesalahan dalam klasifikasi sel darah merah abnormal.

Zhang dkk. [75] Mengusulkan model berdasarkan algoritma pembelajaran mesin pohon keputusan bernama Extreme Gradient Boosting (XGBoost) untuk prediksi waktu merokok reguler. Selanjutnya, untuk membuat kumpulan data simulasi untuk data waktu merokok, Pusat Pengendalian dan Pencegahan Penyakit China mengumpulkan informasi orang dari perokok. Juga, mereka menggunakan modul untuk mengekstrak informasi fitur. Untuk melihat outputnya dalam modul ekstraksi fitur, mereka menggunakan modul pohon keputusan (XGBoost) dan algoritma machine learning Random Forest. Hasil penelitian menunjukkan bahwa efisiensi DT lebih tinggi dari RF, mencapai 84,11 % dengan akurasi DT, sedangkan 58,11 % dengan akurasi RF.

Nandhini dan KS [76] membahas metode efektif untuk mengembangkan model pembelajaran mesin menggunakan beberapa algoritma umum yang dapat membedakan apakah email adalah spam atau ham. Toko Pembelajaran Mesin UCI digunakan sebagai kumpulan data untuk Spambase. Selain itu, mereka mengevaluasi keluaran Logistic Regression (LR), DT, Naïve Bayes (NB), KNN, dan Support Vector Machine (SVM) untuk membangun model pembelajaran mesin yang efisien untuk spam. Menggunakan alat Weka untuk melatih dan mengevaluasi pengumpulan data. Hasil penelitian menunjukkan bahwa kinerja DT sebanding dan lebih baik dari kinerja KNN, dan akurasi keduanya adalah sebagai berikut: DT 99,93 persen, KNN 99,93 persen, LR 93,13 persen, SVM 90,76 persen, dan NB 79,52 persen. .

Talaba dan Ismail [77] mengembangkan pendekatan pembelajaran mesin baru untuk pohon keputusan hibrida dan algoritma genetik yang dikenal sebagai GADT untuk deteksi spam. Algoritma yang paling signifikan untuk meningkatkan efisiensi pohon keputusan adalah algoritma genetika. Juga, ini efisien dan dapat diandalkan untuk klasifikasi teks. Sebuah algoritma genetika telah menggunakan elemen kepercayaan yang mengatur pemangkasan pohon keputusan untuk mengoptimalkan dan mendeteksi nilai optimalnya. Mereka menggunakan set data spam UCI Machine Learning Store. Selain itu, mereka menggunakan mekanisme Prinsip Komponen Analisis (PCA) utama untuk menghapus fitur yang tidak sesuai untuk konten pesan email dan memprosesnya lebih jarang. Temuan menunjukkan bahwa setelah menggunakan PCA, pendekatan GADT campuran memiliki akurasi 93,4% sebelum menggunakan PCA dan akurasi 95,5%. Ini menyiratkan bahwa ekstraksi karakteristik yang tidak sesuai memiliki dampak besar pada PCA.

Arowolo dkk. [78] menerapkan algoritma ekstraksi fitur Analisis Komponen Prinsip (PCA) untuk mengurangi dimensi dan menunjukkan bukti analisis dimensi tinggi pada ekspresi gen. Klasifikasi KNN dan algoritma DT digunakan untuk mendeteksi berbagai struktur biologis dan untuk menawarkan resolusi nilai yang lebih baik serta untuk mendeteksi gen malaria baru dan tes prediksi. Sequencing asam ribonukleat (RNA-seq) juga digunakan sebagai pengumpulan data. Hasil penelitian menunjukkan bahwa kinerja klasifikasi KNN lebih baik daripada klasifikasi DT pada ekstraksi ciri PCA. Akurasi KNN mencapai 86,7% sedangkan DT mencapai 83,3%.

Pathan dkk. [79] mengusulkan teknik baru yang mengenali dan menghilangkan arteri darah untuk segmentasi yang benar dari Optic Disc (OD). Ini dilakukan dengan dua cara. Pertama, filter arah digunakan untuk membangun algoritma identifikasi dan eksklusi pembuluh darah yang efisien. Pada langkah kedua, untuk mendeteksi kontur cakram optik, pengklasifikasi pohon keputusan digunakan untuk mencapai ambang adaptif. Selain itu, dua database terpisah digunakan, termasuk 300 gambar fundus yang diperoleh dari Kasturba Medical College (KMC) Manipal dan juga database RIM-ONE yang dapat diakses publik. Hasil penelitian menunjukkan bahwa teknik segmentasi OD sepenuhnya otomatis yang menggunakan pengklasifikasi pohon keputusan untuk mencapai ambang segmentasi meningkatkan ketahanan algoritma bahkan untuk gambar yang mengandung eksudat, atrofi vesikel, dan pembalikan, Oleh karena itu, menghasilkan fraksinasi OD yang sesuai. Studi matematika menunjukkan efek dari pretreatment. Oleh karena itu, nilai rata-rata akurasi

24

*Tijo & Abdulazeez / Journal of Applied Science and Technology Trends Vol. 02, No. 01, pp. 20 – 28 (2021)*

diperoleh untuk citra KMC sebesar 99,61% dan untuk database RIM-ONE diperoleh nilai rata-rata akurasi sebesar 99,15%.

Nagra dkk. [80] memperkenalkan Self-Inertia Weight Adaptive Particle Swarm Optimization dengan Gradient Base Local Search (SIW-APSO-LS) pendekatan pemilihan fitur dimodifikasi untuk melakukan seleksi fitur dan metode pohon keputusan C4.5 digunakan sebagai pengklasifikasi untuk menentukan sub-set fitur yang diberikan. Saat membandingkan algoritme dalam masalah pemilihan fitur, 16 set data dari UCI Machine Learning Repository digunakan untuk eksperimen. Hasil eksperimen menunjukkan bahwa SIW-APSO-LS menyederhanakan pengumpulan fitur dengan secara efektif mengurangi jumlah fitur yang dipilih, sehingga mempertahankan presisi terbaik dibandingkan dengan pendekatan pemilihan literatur lain untuk fungsi pengujian yang sama. Di bidang pengumpulan atribut, temuan eksperimental menunjukkan bahwa pendekatan yang diusulkan bermanfaat dan akurasi tertinggi yang diperoleh dari total 16 kumpulan data adalah 99,88%

Ahmim et al. [81] mengusulkan Intrusion Detection System (IDS) baru yang menggabungkan beragam sistem klasifikasi yang berbasis DT dan konsep berbasis aturan, yaitu pohon REP, algoritma JRip, dan Forest PA. Secara khusus, pendekatan pertama dan kedua mengambil fitur kumpulan data sebagai input dan mengkategorikan lalu lintas jaringan sebagai Serangan/jinak. Dibandingkan dengan hasil pengklasifikasi pertama dan kedua untuk referensi, pengklasifikasi ketiga menggunakan atribut dari koleksi data asli. Temuan penelitian yang dicapai dengan menggunakan dataset CICIDS2017 untuk menganalisis IDS membuktikan dominasinya dalam hal akurasi, tingkat identifikasi, tingkat alarm palsu, dan overhead waktu relatif terhadap skema mutakhir saat ini. Secara menyeluruh, dengan 94,457%, model kami memiliki DR tertinggi, presisi tertinggi dengan 96,665%, dan FAR terendah dengan 1,145%, meskipun waktu komputasinya yang rendah membuatnya cepat diimplementasikan ke dalam sistem waktu nyata yang lembut.

Li dkk. [82] menyediakan pohon keputusan pembuktian untuk mengklasifikasikan kumpulan data fuzzy dan entropi ding telah digunakan sebagai indikator aturan partisi untuk konstruksinya. Selain itu, Basic Belief Assignment (BBA) dari Iris dan Wine Dataset digunakan untuk menghitung fitur pemisahan yang optimal. Semakin rendah entropi Deng, semakin efektif fitur tersebut untuk mengkarakterisasi sampel. Berbeda dengan aturan campuran standar yang digunakan untuk kombinasi BBA, bukti DT dapat diperluas secara khusus untuk klasifikasi. Temuan menunjukkan bahwa penerapan pembuktian DT berdasarkan entropi keyakinan efektif menurunkan kompleksitas klasifikasi data fuzzy apakah pasien dipengaruhi oleh jenis kanker Malign atau Jinak. Dataset Kanker Payudara Wisconsin, yang berisi 32 atribut dan 569 data, digunakan. Mereka menggunakan uji validasi silang 10 kali lipat untuk mengidentifikasi dan menganalisis algoritma. Akurasinya adalah 95% saat menggunakan dataset Wine, tetapi akurasi yang didapat oleh dataset Iris adalah 98%.

Sathiyanarayanan dkk. [83] menggunakan algoritma DT di bawah mekanisme pembelajaran yang diawasi untuk mengungkap kanker payudara. Identifikasi kanker payudara dilakukan di sini dan difokuskan pada data, yang memisahkan data untuk proses persiapan dan pengujian. Hasil yang diperoleh kontras antara algoritma KNN dan DT. Hasil temuan menunjukkan bahwa akurasi yang

diperoleh KNN adalah 97%, sedangkan DT mencapai akurasi maksimum 99%. Oleh karena itu, algoritma pohon keputusan yang berada di bawah metode pembelajaran yang diawasi memprediksi jenis kanker.

IV. PERBANDINGANDAN PEMBAHASAN Algoritmaklasifikasi

pohon keputusan terdiri dari beberapa jenis yang digunakan untuk membangkitkan DT. Ini adalah dengan mengontrol atribut kontinu dan periodik dari nilai-nilai yang hilang. DT dihasilkan oleh formulir yang biasanya direpresentasikan sebagai pengklasifikasi statistik dan dapat digunakan untuk pengelompokan. Node dan cabang termasuk dalam DT. Setiap node membutuhkan masalah yang didasarkan pada satu atau lebih properti, yaitu membandingkan nilai atribut dengan konstanta atau menggunakan fungsi lain untuk membandingkan lebih dari satu properti. Untuk tujuan pohon keputusan, pengumpulan data pembelajaran kadang-kadang disebut sebagai pohon hasil. Untuk menggabungkan klasifikasi dalam pembelajaran mesin dan penambangan data menggunakan algoritma DT. Dalam urutan berikut, algoritma ini diterapkan secara iteratif dan klasifikasi memerlukan proses tiga tahap: Model Konstruk (Pembelajaran), Model Evaluasi (Akurasi), dan Penggunaan Model (Klasifikasi). Tahap klasifikasi DT didasarkan pada persentase informasi yang diperoleh yang diukur dengan entropi. Metrik jangkauan digunakan untuk menggambarkan karakteristik pengujian untuk sebuah simpul di pohon dan disebut sebagai skala pemilihan properti (properti). Sebagai fungsi uji untuk node saat ini, properti pengetahuan terbaik dihitung. Beberapa studi mengusulkan pendekatan untuk mengatasi kekurangan masalah DT sehingga pohon yang optimal dapat dihitung, berdasarkan tinjauan yang dilakukan sebelumnya, tanpa rincian rinci dan sampel. Metode DT telah menunjukkan bahwa masalah seperti yang dijelaskan di atas dapat dihindari. Selanjutnya, itu akan memberikan dataset yang ditentukan dengan solusi yang tepat. Menurut "Tabel III" diamati bahwa dalam banyak penelitian dilakukan dengan kumpulan data yang berbeda dan pendekatan DT digunakan untuk mengatasi kelemahannya dan untuk mendapatkan kinerja yang lebih baik. Beberapa teknik optimasi telah digunakan dalam penelitian [76] untuk memperkuat pohon keputusan pada dataset ML UCI yang disimpan; Berdasarkan hasil asesmen menunjukkan bahwa pendekatan DT mendapatkan akurasi tertinggi yaitu 99,93% dibandingkan dengan teknik lain seperti KNN, LR, SVM, dan NB yang kinerjanya kurang dari pendekatan DT. Dalam tugas segmentasi, penelitian [79] menggunakan pendekatan DT untuk mengidentifikasi dan mengekstrak arteri darah untuk segmentasi Optic Disc (OD) yang tepat, yang menghasilkan hasil yang lebih besar sebesar 99,61%. Selain itu, berdasarkan penelitian [69], telah ditunjukkan bahwa metode DST juga dapat meningkatkan DT, di mana keduanya menggunakan PT dan MLT untuk DT di dataset UCI; telah ditunjukkan bahwa DST lebih mampu meningkatkan DT daripada teknik lainnya. Pada akhirnya, dengan menggunakan set data UCI Machine Learning Library dan set data CICIDS2017 yang terdiri dari serangan terbaru di antara semua set data lainnya, DT terbukti menjadi yang tertinggi, dan akurasinya memiliki performa terbaik. Sementara penelitian [75] menggunakan DT(XGBoost) dan RF pada dataset Perokok dari Pusat Pengendalian dan Pencegahan Penyakit China, ditemukan bahwa, sekali lagi, pendekatan DT mencapai akurasi tertinggi; yaitu 84,11%. Selanjutnya, berdasarkan penelitian [73], [78], [83] menggunakan DT dan KNN dalam dataset CICIDS2017, RNA-seq Malaria, dan Wisconsin Breast Cancer, ditemukan bahwa DT

25

*Tijo & Abdulazeez / Journal of Applied Science dan Tren Teknologi Vol. 02, No. 01, hlm. 20 – 28 (2021)*

pendekatan memiliki akurasi tertinggi di ketiga studi. Selain itu, akurasinya lebih tinggi saat menggunakan dataset CICIDS2017 yang mencapai akurasi 99,91%.

V.KESIMPULAN

Pengklasifikasi pohon keputusan dikenal karena pandangan mereka yang ditingkatkan tentang hasil kinerja. Karena presisi yang kuat, parameter pemisahan yang dioptimalkan, dan teknik pemangkasan pohon yang disempurnakan (ID3, C4.5, CART, CHAID, dan QUEST) biasanya digunakan oleh semua pengklasifikasi data yang dikenal. Kumpulan data terpisah digunakan untuk melatih sampel dari kumpulan data besar, yang pada gilirannya mempengaruhi presisi kumpulan uji. Pohon keputusan memiliki beberapa kemungkinan kekhawatiran tentang ketahanan, adaptasi skalabilitas dan optimalisasi ketinggian. Namun, berbeda dengan metode klasifikasi data lainnya, pohon keputusan membuat kumpulan aturan yang efisien yang mudah dipahami. Makalah ini mengulas penelitian terbaru yang dilakukan di banyak bidang, seperti analisis penyakit medis, klasifikasi teks, klasifikasi smartphone dan gambar pengguna, dll. Selanjutnya, detail yang digunakan dalam teknik/algoritma, kumpulan data digunakan oleh penulis dan hasil yang dicapai terkait dengan akurasi dirangkum untuk pohon keputusan. Terakhir, akurasi terbaik yang dicapai untuk algoritma pohon keputusan adalah 99,93% ketika menggunakan repositori pembelajaran mesin sebagai kumpulan data.

[REFERENSI

1] D. Abdulqader, A. Mohsin Abdulazeez, dan D. Zeebaree, “Machine Learning Supervised Algorithms of Gene Selection: A Review,” Apr. 2020.

[2] MW Libbrecht dan WS Noble, “Aplikasi pembelajaran mesin di genetika dan genomik,” *Nature Review Genetics*, vol. 16, tidak. 6, pp. 321–332, 2015.

[3] J. Wang, P. Neskovic, dan LN Cooper, “Pelatihan Seleksi Data untuk Mesin Vektor Dukungan,” di *Advances in Natural Computation*, vol. 3610, L. Wang, K. Chen, dan YS Ong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, hlm. 554–564.

[4] D. Maulud dan AM Abdulazeez, “Sebuah Tinjauan tentang Regresi Linier Komprehensif dalam Pembelajaran Mesin,” *Jurnal Tren Sains dan Teknologi Terapan*, vol. 1, tidak. 4, hlm. 140–147, 2020.

[5] G. Carleo *dkk.*, "Pembelajaran mesin dan ilmu fisika," *Ulasan Fisika Modern*, vol. 91, tidak. 4, hal. 045002, 2019. [6] T. Hillel, M. Bierlaire, M. Elshafie, dan Y. Jin, “Sebuah tinjauan sistematis metodologi klasifikasi pembelajaran mesin untuk pemodelan pilihan mode penumpang,” *Journal of Choice Modeling*, p. 100221, 2020.

[7] D. Zeebaree, H. Haron, A. Mohsin Abdulazeez, dan D. Zebari, *Machine learning and Region Growing for Breast Cancer Segmentation*. 2019, hal. 93.

[8] C. Feng, S. Wu, dan N. Liu, “Sebuah kerangka pembelajaran mesin yang berpusat pada pengguna untuk pusat operasi keamanan siber,” pada *Konferensi Internasional IEEE 2017 tentang Intelijen dan Keamanan Informatika (ISI)*, Beijing, Cina , Juli 2017, hlm. 173–175, doi: 10.1109/ISI.2017.8004902.

[9] SB Kotsiantis, I. Zaharakis, dan P. Pintelas, "Pembelajaran mesin yang diawasi: Tinjauan teknik klasifikasi," *Muncul aplikasi kecerdasan buatan dalam teknik komputer*, vol. 160, tidak. 1, pp. 3–24, 2007.

[10] SB Kotsiantis, ID Zaharakis, dan PE Pintelas, “Pembelajaran mesin: tinjauan teknik klasifikasi dan penggabungan,” *Artif Intell Rev*, vol. 26, tidak. 3, hlm. 159–190, November 2006, doi: 10.1007/s10462-007-9052-3.

[11] C. Surv, MN Murty, PJ Flynn, AK Jain, dan PJ Flynn, *Dan*. 1999.

[12] D. Sharma dan N. Kumar, “A Review on Machine Learning Algorithms, Tasks and Applications,” vol. 6, pp. 2278–1323, Oktober 2017.

[13] K. Pahwa dan N. Agarwal, “Stock Market Analysis using Supervised Machine Learning,” pada *Konferensi Internasional 2019 tentang Pembelajaran Mesin, Data Besar, Cloud dan Komputasi Paralel (COMITCon )*, Faridabad, India, Februari 2019, hlm. 197–200, doi: 10.1109/COMITCon.2019.8862225.

[14] M. Pérez-Ortiz, S. Jiménez-Fernández, PA Gutiérrez, E. Alexandre, C. Hervás-Martínez, dan S. Salcedo-Sanz, “A Review of Classification Problems and Algorithms in Renewable Energy Applications,” *Energies*, vol. 9, tidak. 8, Seni. tidak. 8 Agustus 2016, doi: 10.3390/en9080607.

[15] Anuradha and G. Gupta, “A self explanatory review of decision tree classifiers,” dalam *Konferensi Internasional tentang Kemajuan Terbaru dan Inovasi dalam Teknik (ICRAIE-2014)*, Jaipur, India, Mei 2014, hlm. 1–7, doi : 10.1109/ICRAIE.2014.6909245.

[16] S. Patil dan U. Kulkarni, “Accuracy Prediction for Distributed Decision Tree using Machine Learning approach,” in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, hlm. 1365–1371, doi : 10.1109/ICOEI.2019.8862580.

[17] NS Ahmed dan MH Sadiq, “Clarify of the random forest algorithm in an education field,” di *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, hlm. 179–184.

[18] D. Zeebaree, *Seleksi Gen dan Klasifikasi Data Microarray Menggunakan Convolutional Neural Network*. 2018.

[19] OM Salih Hassan, A. Mohsin Abdulazeez, dan VM Tiryaki, “Gait Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis,” dalam *International Conference on Advanced Science and Engineering (ICOASE) 2018*, Duhok , Oktober 2018, hlm. 173–178, doi: 10.1109/ICOASE.2018.8548909.

[20] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, dan J. Saeed, "Tinjauan Komprehensif Teknik Pengurangan Dimensi untuk Seleksi Fitur dan Ekstraksi Fitur," *Jurnal Tren Sains dan Teknologi Terapan*, vol. 1, tidak. 2, pp. 56–70, 2020.

[21] DV Patil dan RS Bichkar, “A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets,” pada *Konferensi Internasional IEEE 2006 tentang Teknologi Industri*, Desember 2006, hlm. 429–433, doi: 10.1109/ICIT.2006.372250.

[22] O. Ahmed dan A. Brifcani, “Klasifikasi Ekspresi Gen Berdasarkan Pembelajaran Mendalam,” dalam *Konferensi Internasional Ilmiah ke-4 Najaf (SICN) 2019 ke*, Al-Najef, Irak, April 2019, hlm. 145–149, doi: 10.1109/SICN47020.2019.9019357.

[23] MA Sulaiman, “Evaluating Data Mining Classification Methods Performance in Internet of Things Applications,” *Journal of Soft Computing and Data Mining*, vol. 1, tidak. 2, hlm. 11–25, 2020.

[24] F. Yang, “An Extended Idea about Decision Trees,” dalam *Konferensi Internasional 2019 tentang Ilmu Komputasi dan Kecerdasan Komputasi (CSCI)*, Desember 2019, hlm. 349–354, doi: 10.1109/CSCI49370.2019.00068.

[25] J. Liang, Z. Qin, S. Xiao, L. Ou, dan X. Lin, “Klasifikasi pohon keputusan yang efisien dan aman untuk layanan diagnosis online berbantuan cloud,” *Transaksi IEEE pada Komputasi yang Dapat Diandalkan dan Aman*, 2019.

[26] A. Mohsin Abdulazeez, A. Brifcani, dan Issa, “Deteksi Intrusi dan Pengklasifikasi Serangan Berdasarkan Tiga Teknik: Studi Perbandingan Deteksi Intrusi dan Pengklasifikasi Serangan Berdasarkan Tiga Teknik: Studi Perbandingan 387,” Januari 2021.

[27] AS Eesa, Z. Orman, dan AMA Bricfani, “Pendekatan pemilihan fitur baru berdasarkan algoritma optimisasi sotong untuk sistem deteksi intrusi,” *Sistem Pakar dengan Aplikasi*, vol. 42, tidak. 5, hlm. 2670–2679, April 2015, doi: 10.1016/j.eswa.2014.11.009.

[28] A. Shamim, H. Hussain, dan Maqbool Uddin Shaikh, “A framework for generation of rules from decision tree and decision table,” dalam *Konferensi Internasional 2010 tentang Teknologi Informasi dan Berkembang*, Juni 2010, hlm. 1–6 , doi: 10.1109/ICIET.2010.5625700.

[29] A. Suresh, R. Udendhran, dan M. Balamurgan, "Jaringan saraf hibrida dan pengklasifikasi berbasis pohon keputusan untuk pengambilan keputusan prognostik pada kanker payudara," *Soft Computing*, vol. 24, tidak. 11, hlm. 7947– 7953, 2020.

26

*Tijo & Abdulazeez / Jurnal Tren Sains dan Teknologi Terapan Vol. 02, No. 01, hlm. 20 – 28 (2021)*

[30] Priyanka dan D. Kumar, “Pengklasifikasi pohon keputusan: survei terperinci,” *Jurnal Internasional Ilmu Informasi dan Keputusan*, vol. 12, tidak. 3, pp. 246–269, 2020.

[31] AS Eesa, AM Abdulazeez, dan Z. Orman, “A DIDS Based on The Combination of Cuttlefish Algorithm and Decision Tree,” *Science Journal of University of Zakho*, vol. 5, tidak. 4, pp. 313–318, 2017.

[32] R. Kumar dan R. Verma, “Algoritma klasifikasi untuk penambangan data: Sebuah survei,” *Jurnal Internasional Inovasi dalam Rekayasa dan Teknologi (IJIET)*, vol. 1, tidak. 2, pp. 7–14, 2012.

[33] SS Nikam, “Sebuah studi perbandingan teknik klasifikasi dalam algoritma data mining,” *Oriental journal of computer science & technology*, vol. 8, tidak. 1, hlm. 13–19, 2015.

[34] CZ Janikow, “Pohon keputusan fuzzy: masalah dan metode,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, tidak. 1, hlm. 1–14, 1998.

[35] G. Stein, B. Chen, AS Wu, dan KA Hua, "Pengklasifikasi pohon keputusan untuk deteksi intrusi jaringan dengan pemilihan fitur berbasis GA," dalam *Prosiding tahunan ke-43 Konferensi regional Tenggara Jilid 2*, 2005, hlm. 136–141.

[36] IS Damanik, AP Windarto, A. Wanto, SR Andani, dan W. Saputra, “Optimalisasi Pohon Keputusan pada C4. 5 Algoritma Menggunakan Algoritma Genetika,” dalam *Journal of Physics: Conference Series*, 2019, vol. 1255, tidak. 1, hal. 012012.

[37] R. Barros, M. Basgalupp, A. de Carvalho, dan A. Freitas, “A Survey of Evolutionary Algorithms for Decision-Tree Induction,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications dan Ulasan*, vol. 42, hlm. 291–312, Januari 2012, doi: 10.1109/TSMCC.2011.2157494.

[38] G. Gupta, "Sebuah tinjauan penjelasan diri pengklasifikasi pohon keputusan," dalam *konferensi Internasional tentang kemajuan terbaru dan inovasi dalam rekayasa (ICRAIE-2014)*, 2014, hlm. 1-7.

[39] SS Gavankar dan SD Sawarkar, “Pohon keputusan yang bersemangat,” pada *Konferensi Internasional ke-2 untuk Konvergensi Teknologi (I2CT)*2017 , Mumbai, April 2017, hlm. 837–840, doi: 10.1109/I2CT.2017.8226246.

[40] PH Swain dan H. Hauska, “The decision tree classifier: Design and potential,” *IEEE Transactions on Geoscience Electronics*, vol. 15, tidak. 3, pp. 142-147, 1977.

[41] A. Dey, "Algoritma pembelajaran mesin: tinjauan," *Jurnal Internasional Ilmu Komputer dan Teknologi Informasi*, vol. 7, tidak. 3, hlm. 1174–1179, 2016.

[42] J. Mrva, . Neupauer, L. Hudec, J. evcech, dan P. Kapec, “Dukungan Keputusan dalam Data Medis Menggunakan Visualisasi Pohon Keputusan 3D,” dalam *Konferensi E-Health dan Bioengineering (EHB)*, November 2019, hlm. 1-4, doi: 10.1109/EHB47216.2019.8969926.

[43] Y. Bengio, O. Delalleau, dan C. Simard, "POHON KEPUTUSAN TIDAK UMUM UNTUK VARIASI BARU," *KECERDASAN KOMPUTASI*, hlm. 19.

[44] CE Brodley dan PE Utgoff, “Pohon keputusan multivariat,” *Pembelajaran mesin*, vol. 19, tidak. 1, pp. 45-77, 1995.

[45] GKF Tso dan KKW Yau, "Memprediksi konsumsi energi listrik: Perbandingan analisis regresi, pohon keputusan dan jaringan saraf," *Energi*, vol. 32, tidak. 9, hlm. 1761–1768, Sep. 2007, doi: 10.1016/j.energy.2006.11.010.

[46] S. Singh dan P. Gupta, “Studi banding ID3, gerobak dan C4. 5 algoritma pohon keputusan: survei,” *International Journal of Advanced Information Science and Technology (IJAIST)*, vol. 27, tidak. 27, pp. 97– 103, 2014.

[47] L. Rokach dan O. Maimon, “Top-Down Induction of Decision Trees Classifiers—A Survey,” *Systems, Man, and Cybernetics, Part C: Applications and Review, IEEE Transaksi pada*, vol. 35, hlm. 476– 487, Desember 2005, doi: 10.1109/TSMCC.2004.843247.

[48] ​​T.-S. Lim, W.-Y. Loh, dan Y.-S. Shih, “Perbandingan akurasi prediksi, kompleksitas, dan waktu pelatihan dari tiga puluh tiga algoritma klasifikasi lama dan baru,” *Machine learning*, vol. 40, tidak. 3, hlm. 203– 228, 2000.

[49] W.-Y. Loh, "Lima Puluh Tahun Pohon Klasifikasi dan Regresi," *Tinjauan Statistik Internasional*, vol. 82, Juni 2014, doi: 10.1111/insr.12016.

[50] SR Jiao, J. Song, dan B. Liu, “A Review of Decision Tree Classification Algorithms for Continuous Variables,” dalam *Journal of Physics: Conference Series*, 2020, vol. 1651, tidak. 1, hal. 012083.

[51] Y.-Y. Song dan Y. Lu, “Metode pohon keputusan: aplikasi untuk klasifikasi dan prediksi,” *Arsip psikiatri Shanghai*, vol. 27, hlm. 130–5, April 2015, doi: 10.11919/j.issn.1002-0829.215044.

[52] RekhaMolala, “Entropi, Perolehan Informasi, dan Indeks Gini; inti dari Pohon Keputusan,” *Medium*, 23 Maret 2020. https://blog.clairvoyantsoft.com/entropy-information-gain-and-gini index-the-crux-of-a-decision-tree-99d0cdc699f4 (diakses 28 Desember 2020).

[53] V. Cheushev, DA Simovici, V. Shmerko, dan S. Yanushkevich, "Entropi fungsional dan pohon keputusan," dalam *Prosiding. Simposium Internasional IEEE ke-28 tahun 1998 tentang Logika Bernilai Ganda (Kat. No. 98CB36138)*, 1998, hlm. 257–262.

[54] X. Chen, Z. Yang, dan W. Lou, “Diagnosis Kesalahan Rolling Bearing Berdasarkan Entropi Permutasi VMD dan Pohon Keputusan,” pada *Konferensi Internasional ke-3 2019 tentang Teknologi Informasi Elektronik dan Teknik Komputer (EITCE)*, Xiamen, Tiongkok, Oktober 2019, hlm. 1911–1915, doi: 10.1109/EITCE47263.2019.9095187.

[55] C. Shang, M. Li, S. Feng, Q. Jiang, dan J. Fan, "Pemilihan fitur melalui memaksimalkan perolehan informasi global untuk klasifikasi teks," *Sistem Berbasis Pengetahuan*, vol. 54, hlm. 298–309, Des. 2013, doi: 10.1016/j.knosys.2013.09.019.

[56] T. Maszczyk dan W. Duch, “Perbandingan entropi Shannon, Renyi dan Tsallis yang digunakan dalam pohon keputusan,” dalam *Konferensi Internasional tentang Kecerdasan Buatan dan Komputasi Lunak*, 2008, hlm. 643–651.

[57] LE Raileanu dan K. Stoffel, "Perbandingan Teoritis antara Indeks Gini dan Kriteria Keuntungan Informasi," *Annals of Mathematics and Artificial Intelligence*, vol. 41, tidak. 1, hlm. 77–93, Mei 2004, doi: 10.1023/B:AMAI.0000018580.96245.c6.

[58] Y. Liu, L. Hu, F. Yan, dan B. Zhang, “Information Gain with Weight Based Decision Tree for the Employment Forecasting of Undergraduates,” pada *2013 IEEE International Conference on Green Computing and Communications dan IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Beijing, China, Agustus 2013, hlm. 2210–2213, doi: 10.1109/GreenCom-iThings CPSCom.2013.417.

[59] RL De Mántaras, “Pengukuran pemilihan atribut berbasis jarak untuk induksi pohon keputusan,” *Machine learning*, vol. 6, tidak. 1, hlm. 81–92, 1991.

[60] S. Taneja, C. Gupta, K. Goyal, dan D. Gureja, “Algoritme k-nearest tetangga yang disempurnakan menggunakan perolehan informasi dan pengelompokan,” pada *Konferensi Internasional Keempat 2014 tentang Teknologi Komputasi & Komunikasi Tingkat Lanjut*, 2014, hlm. 325–329.

[61] Y. Zhao dan Y. Zhang, "Perbandingan metode pohon keputusan untuk menemukan objek aktif," *Advances in Space Research*, vol. 41, tidak. 12, pp. 1955–1959, 2008.

[62] K. Mittal, D. Khanduja, dan PC Tewari, “Sebuah wawasan tentang 'Analisis Pohon Keputusan'”,” *World Wide Journal of Multidisipliner Penelitian dan Pengembangan*, vol. 3, tidak. 12, hlm. 111–115, 2017.

[63] Priyanka dan D. Kumar, “Pengklasifikasi pohon keputusan: survei terperinci,” *Jurnal Internasional Ilmu Informasi dan Keputusan*, vol. 12, tidak. 3, pp. 246–269, 2020.

[64] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, dan H. Tang, “Memprediksi diabetes mellitus dengan teknik pembelajaran mesin,” *Perbatasan dalam genetika*, vol. 9, hal. 515, 2018.

[65] TA Assegie dan PS Nair, “Pengenalan angka tulisan tangan dengan klasifikasi pohon keputusan: pendekatan pembelajaran mesin,” *Jurnal Internasional Teknik Elektro dan Komputer*, vol. 9, tidak. 5, hal. 4446, 2019.

[66] F. De Felice *dkk.*, “Algoritme pohon keputusan pada kanker rektum stadium lanjut lokal: contoh interpretasi berlebihan dan penyalahgunaan pendekatan pembelajaran mesin,” *Journal of Cancer Research and Clinical Oncology*, vol. 146, tidak. 3, pp. 761–765, 2020.

[67] IH Sarker, A. Colman, J. Han, AI Khan, YB Abushark, dan K. Salah, “Behavdt: a behavioral decision tree learning to build user centric context-aware model prediktif,” *Jaringan dan Aplikasi Seluler*, vol. 25, tidak. 3, pp. 1151-1161, 2020.

[68] X. Hu, C. Rudin, dan M. Seltzer, “Pohon keputusan sparse yang optimal,” dalam *Advances in Neural Information Processing Systems*, 2019, hlm. 7267– 7275.

[ 69] S. Patil dan U. Kulkarni, “Prediksi Akurasi untuk Pohon Keputusan Terdistribusi menggunakan pendekatan Pembelajaran Mesin,” pada *Konferensi Internasional ke-3 tentang Tren Elektronika dan Informatika*

27

*Tijo & Abdulazeez / Jurnal Tren Sains dan Teknologi Terapan Vol. 02, No. 01, hlm. 20 – 28 (2021)*

*(ICOEI)*, April 2019, hlm. 1365–1371, doi: 10.1109/ICOEI.2019.8862580.

[70] D. Hussain, MA Al-Antari, MA Al-Masni, S.-M. Han, dan T.-S. Kim, "Segmentasi tulang paha dalam pencitraan DXA menggunakan pohon keputusan pembelajaran mesin," *Journal of X-ray Science and Technology*, vol. 26, tidak. 5, pp. 727–746, 2018.

[71] N. Linty, A. Farasin, A. Favenza, dan F. Dovis, “Detection of GNSS Ionosphere Scintillations Based on Machine Learning Decision Tree,” *IEEE Transactions on Aerospace and Electronic Sistem*, vol. 55, tidak. 1, hlm. 303–317, Februari 2019, doi: 10.1109/TAES.2018.2850385.

[72] W. Kuang, Y. Chan, S. Tsang, dan W. Siu, “Keputusan Mode Intra Cepat Berbasis Pembelajaran Mesin untuk Pengkodean Konten Layar HEVC melalui Pohon Keputusan,” *Transaksi IEEE pada Sirkuit dan Sistem untuk Teknologi Video*, vol . 30, tidak. 5, hlm. 1481–1496, Mei 2020, doi: 10.1109/TCSVT.2019.2903547.

[73] I. Ramadhan, P. Sukarno, dan MA Nugroho, “Analisis Perbandingan K-Nearest Neighbor dan Pohon Keputusan dalam Mendeteksi Distributed Denial of Service,” pada *Konferensi Internasional ke-8 tahun 2020 tentang Teknologi Informasi dan Komunikasi (ICoICT)*, Yogyakarta, Indonesia, Juni 2020, hlm. 1-4, doi: 10.1109/ICoICT49345.2020.9166380.

[74] VME Batitis, MJG Caballes, AA Ciudad, MD Diaz, RD Flores, dan ERE Tolentin, “Klasifikasi Gambar Sel Darah Merah Abnormal Menggunakan Algoritma Pohon Keputusan,” pada *Konferensi Internasional Keempat tentang Metodologi dan Komunikasi Komputasi (ICCMC) 2020*, Maret 2020, hlm. 498–504, doi: 10.1109/ICCMC48092.2020.ICCMC-00093.

[75] Y. Zhang, J. Liu, Z. Zhang, dan J. Huang, “Prediksi Perilaku Merokok Harian Berdasarkan Algoritma Pembelajaran Mesin Pohon Keputusan,” pada *Konferensi Internasional ke-9 IEEE 2019 tentang Informasi Elektronik dan Komunikasi Darurat (ICEIEC)*, Juli 2019, hlm. 330–333, doi: 10.1109/ICEIEC.2019.8784698.

[76] S. Nandhini dan JMKS, “Evaluasi Kinerja Algoritma Pembelajaran Mesin untuk Deteksi Spam Email,” pada *Konferensi Internasional 2020 tentang Tren yang Muncul dalam*

*Teknologi dan Rekayasa Informasi (ic-ETITE)*, Februari 2020, hlm. 1-4, doi: 10.1109/ic-ETITE47903.2020.312.

[77] AI Taloba dan SSI Ismail, “An Intelligent Hybrid Technique of Decision Tree and Genetic Algorithm for E-Mail Spam Detection,” dalam *Konferensi Internasional Kesembilan 2019 tentang Komputasi Cerdas dan Sistem Informasi (ICICIS)*, Desember 2019, hlm. 99 -104, doi: 10.1109/ICICIS46948.2019.9014756.

[78] MO Arowolo, M. Adebiyi, A. Adebiyi, and O. Okesola, “PCA Model For RNA-Seq Malaria Vector Data Classification Using KNN And Decision Tree Algorithm,” pada *Konferensi Internasional 2020 di bidang Matematika, Teknik Komputer dan Ilmu Komputer (ICMCECS)*, Maret 2020, hlm. 1–8, doi: 10.1109/ICMCECS47690.2020.240881.

[79] S. Pathan, P. Kumar, R. Pai, dan SV Bhandary, "Deteksi otomatis kontur cakram optik dalam gambar fundus menggunakan pengklasifikasi pohon keputusan," *Biocybernetics and Biomedical Engineering*, vol. 40, tidak. 1, hlm. 52–64, 2020.

[80] AA Nagra *dkk.*, “Optimasi swarm partikel adaptif bobot inersia diri hibrida dengan pencarian lokal menggunakan C4. 5 pengklasifikasi pohon keputusan untuk masalah pemilihan fitur,” *Connection Science*, vol. 32, tidak. 1, pp. 16–36, 2020.

[81] A. Ahmim, L. Maglaras, MA Ferrag, M. Derdour, dan H. Janicke, “Sistem deteksi intrusi hierarkis baru berdasarkan pohon keputusan dan model berbasis aturan, ” pada *Konferensi Internasional ke-15 2019 tentang Komputasi Terdistribusi dalam Sistem Sensor (DCOSS)*, 2019, hlm. 228–233.

[82] M. Li, H. Xu, dan Y. Deng, "Pohon keputusan bukti berdasarkan entropi keyakinan," *Entropi*, vol. 21, tidak. 9, hal. 897, 2019.

[83] P. Sathiyanarayanan, S. Pavithra, MS SARANYA, dan M. Makeswari, “Identification of Breast Cancer Using The Decision Tree Algorithm,” dalam *Konferensi Internasional IEEE 2019 tentang Sistem, Komputasi, Otomasi dan Jaringan (ICSCAN )*, 2019, hlm. 1–6.

28