BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Untuk mendukung data-data dan teori-teori yang digunakan di dalam penelitian ini, penulis merujuk beberapa artikel penelitian yang berkaitan dengan masalah penelitian yang diangkat. Tabel II.1 menunjukkan rangkuman dari artikel-artikel penelitian yang digunakan sebagai bahan rujukan di dalam penelitian ini.

Tabel II.1 Penelitian Terkait Algoritma C4.5

Penulis	Penelitian	Hasil Penelitian
(Jinan et al., 2023)	Kombinasi algoritma C4.5 dan VGG-19 dalam klasifikasi ras anjing *bulldog*	Algoritma C4.5 memiliki nilai akurasi sebesar 90%
(M. Z. Lubis et al., 2023)	Penerapan algoritma C4.5 dalam klasifikasi sentimen di media sosial Twitter dengan cakupan masalah pemblokiran PayPal oleh Kominfo	Algoritma C4.5 memiliki nilai akurasi sebesar 97.6%
(Khasanah et al., 2022)	Penerapan algoritma C4.5 dalam klasifikasi data jasa pengiriman pada perusahaan pelayaran	Algoritma C4.5 memiliki tingkat akurasi rata-rata sebesar 84%
(Lintang et al., 2022)	Menganalisis minat siswa SMA untuk melanjutkan pendidikan ke jenjang perguruan tinggi dengan menggunakan algoritma C4.5	Algoritma C4.5 memiliki nilai akurasi, presisi dan recall: 95.3%, 90.8%, 95.3%

(Irfiani et al., 2022)	Penerapan algoritma C4.5 dalam klasifikasi potensi tsunami berdasarkan gempa bumi di indonesia	Algoritma C4.5 memiliki nilai akurasi, sebesar 99.96%

Tabel II.2 Penelitian Terkait *Random Forest*

Penulis	Penelitian	Hasil Penelitian
(Maturo & Verde, 2022)	Menggabungkan teknik *unsupervised* dan *supervised learning* dalam pengklasifikasian data fungsional	Algoritma *random forest* berhasil meningkatkan nilai akurasi hingga mencapai 89.05%.
(Purwanto et al., 2023)	Pemanfaatan algoritma *decision tree* dan *random forest* dalam pemetaan hutan mangrove di Taman Nasional Sembilang	Algoritma *random forest* memiliki nilai akurasi sebesar 99.12%.
(da Silva et al., 2022)	Perbandingan beberapa metode *machine learning* dalam klasifikasi anjungan minyak berbasis citra SAR C-Band	Algoritma *random forest* memiliki nilai rata-rata *accuracy* sebesar 76%.
(Setiawan et al., 2022)	Pemanfaatan algoritma *random forest* dalam memprediksi sentimen pengguna jasa GoFood pada media sosial Twitter	Algoritma *random forest* memiliki nilai akurasi sebesar 98.6%.
(Margolang & Zarlis, 2023)	Pemanfaatan algoritma *K-means clustering* dan *random forest* dalam mengklasifikasikan sentimen masyarakat terhadap penyelenggaraan event MotoGP Mandalika 2022	Algoritma *random forest* memiliki nilai akurasi, presisi dan recall sebesar 99%, 99.01%, 99%

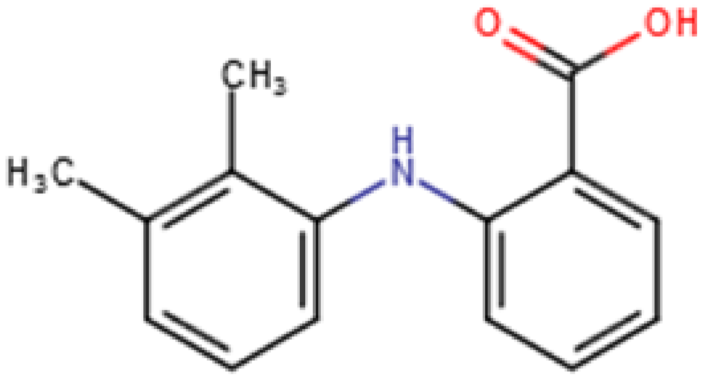
Tabel II.3 Penelitian Terkait C4.5 dan Random Forest

Penulis	Penelitian	Hasil Penelitian
(Rismayati et al., 2022)	Implementasi algoritma *ensemble* seperti C4.5 dan *random forest*, dalam memprediksi tingkat kelulusan mahasiswa	Algoritma C4.5 memiliki nilai akurasi sebesar 83.5%. Algoritma *random forest* memiliki nilai akurasi sebesar 90.4%.
(Singhal et al., 2022)	Membandingkan algoritma C4.5 dan *random forest* dalam memprediksi penyakit kanker payudara	Algoritma C4.5 memiliki nilai akurasi sebesar 95.1%. Algoritma *random forest* memiliki nilai akurasi sebesar 96.5%.
(Riyadi et al., 2023)	Komparasi algoritma C4.5 dan *random forest* dalam memprediksi bakat anak berdasarkan hobi	Algoritma C4.5 memiliki nilai akurasi tertinggi sebesar 92.1%. Algoritma *random forest* memiliki nilai akurasi tertinggi sebesar 92.7%.
(Bhardwaj & Chaurasia, 2022)	Komparasi algoritma C4.5 dan *random forest*, dalam evaluasi seismik likuifaksi tanah	Algoritma C4.5 memiliki nilai akurasi sebesar 97.6%.

		Algoritma *random forest* memiliki nilai akurasi sebesar 98.4%.
(Kapadia et al., 2022)	Analisis hasil pertandingan kriket menggunakan algoritma C4.5 dan *random forest*	Algoritma C4.5 memiliki nilai akurasi sebesar 54%. Algoritma *random forest* memiliki nilai akurasi sebesar 56%.
(Dikananda et al., 2022)	Komparasi algoritma C4.5 dan *random forest* dalam memprediksi tingkat kelulusan mahasiswa	Algoritma C4.5 memiliki akurasi rata-rata sebesar 58.68%. Algoritma *random forest* memiliki akurasi rata-rata sebesar 60.76%.

2.2 Asam Mefenamat (*Mefenamic Acid*)

Asam mefenamat (*mefenamic acid*) merupakan turunan dari asam antranilat, yang tersusun sebagai asam antranilat N-(2,3-xylyl). Asam mefenamat memiliki hubungan yang dengan asam 3-hydroxyanthranilic, metabolit triptofan yang terbentuk secara alami. Gambar II.1 menunjukkan struktur kimia dari asam mefenamat (Srivastava et al., 2019).



Gambar II.1 Struktur Kimia Asam Mefenamat

Sumber: (Srivastava et al., 2019)

Asam mefenamat adalah obat anti inflamasi nonsteroid (*nonsteroidal anti-inflammatory drug*), yang termasuk ke dalam obat kelas BCS II. Obat ini memiliki sifat yang sulit larut dalam air, dengan nilai kisaran dari 0,004 mg/mL pada 37°C hingga 0,2 mg/mL pada 25°C. Namun, obat ini sangat mudah larut dalam pelarut organik hidrofilik, seperti etanol (14,8 mg/mL pada 25°C) dan polietilen glikol 400 (11,5 mg/mL pada 25°C). Sifat kelarutannya di dalam air, yang menentukan laju disolusi asam mefenamat, ini adalah faktor utama dalam ketersediaan hayatinya (Putranti et al., 2019).

Dosis obat yang diperbolehkan pada asam mefenamat adalah antara 500 mg dan 250 mg dengan durasi lebih dari 7 hari. Obat yang termasuk ke dalam asam amino aromatik ini memiliki mekanisme kerja yang menghambat COX (*cyclooxygenase enzymes*), sehingga sangat diperlukan untuk produksi hormon prostaglandin. Senyawa ini memiliki kelarutan yang sangat rendah pada iritasi lambung, cairan biologis dan memiliki pendek waktu aktivitas biologis sekitar dua jam (Mohammed & Kareem, 2020).

2.3 *Machine Learning*

Data mining memiliki keterkaitan dengan bidang-bidang ilmu yang lain, seperti sistem basis data, *data warehousing*, statistik, *machine learning*, dan *information retrieval*. *Data mining* juga didukung oleh bidang-bidang ilmu lain seperti *neural network*, *pattern recognition*, *spatial data analysis*, *image database* dan *signal processing* (Prasetya & Ridwan, 2019).

Machine learning merupakan bagian dari *data mining* yang bertujuan untuk menggali informasi dan pengetahuan dari dalam sebuah *data warehouse*. Data mining adalah kegiatan menemukan pola-pola menarik dari sejumlah besar data yang tersimpan di dalam *database*, *data warehouse* atau penyimpanan informasi lainnya. Berdasarkan kategorinya, *machine learning* dapat dibagi menjadi dua kelompok, yaitu (Suradiradja, 2021):

1. *Supervised Learning*

   Kelompok yang memproses data sudah memiliki sampel masukan dan keluaran yang menjadi target. Dalam *supervised learning*, model hasil pelatihan digunakan untuk memprediksi keluaran dalam bentuk data pengujian.

2. *Unsupervised learning*

   Hampir sama dengan *supervised learning*, namun tidak membedakan antara *training set* dan *test set* dengan data yang tidak berlabel atau target. Data yang dimasukkan pada model ini diproses untuk menemukan pola-pola yang tersembunyi di dalam kumpulan data tersebut.

Machine learning dapat memprediksi pola dan informasi di dalam data dengan tingkat keakuratan yang tinggi, sehingga dapat membantu memecahkan masalah-masalah di dalam kehidupan manusia, khususnya yang berhubungan dengan *big data*. Kemampuan *machine learning* dalam mengolah data dengan jumlah yang sangat besar, bahkan sampai tidak dapat terbayangkan oleh manusia, dapat menghasilkan informasi dalam bentuk keputusan yang cerdas, yang sangat berguna jika diimplementasikan ke dalam berbagai bidang (Abijono et al., 2021).

2.4 Klasifikasi

Klasifikasi merupakan sebuah proses untuk menemukan model atau fungsi yang dapat menjelaskan atau membedakan konsep atau kelas data. Tujuan dari klasifikasi adalah untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui dan memprediksi target kelas untuk setiap kasus dalam data. Sebuah tugas klasifikasi dimulai dengan satu set data di mana kelas dikenal. Dalam proses pelatihannya, algoritma klasifikasi berusaha untuk menemukan hubungan antara nilai-nilai prediksi dengan nilai-nilai target dengan menggunakan teknik yang berbeda. Hasil proses inilah yang kemudian menghasilkan sebuah model, yang kemudian dapat diterapkan pada data yang berbeda (Esthiningtyas et al., 2020).

Sebagai salah satu teknik dalam *machine learning*, klasifikasi dapat digunakan untuk memprediksi keanggotaan kelompok (kelas) dari sekumpulan data, sehingga dapat diperoleh hasil analisis dan deskripsi penting dari kelompok-kelompok tersebut. Sehingga dapat disimpulkan bahwa inti dari klasifikasi adalah bagaimana memisahkan sekumpulan data ke dalam beberapa kelas yang berbeda (Kurniawan et al., 2020).

Beberapa algoritma klasifikasi yang populer dan banyak digunakan di dalam penelitian antara lain: *decision/classification trees*, *bayesian classifiers*/ *naïve bayes classifiers*, *neural networks*, *analisa statistik*, algoritma genetika, *rough sets*, *k-nearest neighbor*, *metode rule based*, *memory based reasoning*, dan *support vector machines* (SVM) (Annur, 2018).

Algoritma klasifikasi dapat diukur dengan menggunakan beberapa parameter seperti akurasi, presisi, *recall* dan *f-measure* untuk menganalisa performanya. Selain itu, parameter lain seperti *kappa statistic*, *mean absolute error* (MAE), *root mean squared error* (RMSE), dan *receiver operating characteristic* (ROC) *Area*, dapat digunakan sebagai ukuran metrik untuk mengukur seberapa baik model klasifikasi dalam memprediksi pola data (Primajaya & Sari, 2018).

2.5 Algoritma C4.5

Algoritma C4.5 merupakan algoritma *data mining* yang tergolong ke dalam *decision tree*, yang dapat digunakan baik untuk masalah klasifikasi maupun masalah prediksi. *Decision tree* merupakan teknik yang memanfaatkan *divide and conquer*, dimana dalam memecahkan masalah terlebih dahulu dibentuk himpunan-himpunan masalah dari satu ruang masalah secara keseluruhan. Pada *decision tree*, proses pemecahan masalah dilakukan dengan cara mengubah data menjadi bentuk sebuah *tree*, yang menghasilkan aturan-aturan (*rule*) yang sudah disederhanakan dalam bentuk parameter-parameter pemecahan masalah klasifikasi maupun prediksi (Jumiyati & Prasetyaningrum, 2020; Ramayu et al., 2022).

C4.5 memiliki kelebihan dalam hal kemudahan interpretasi modelnya yang mudah dan dapat diimplementasikan untuk nilai kontinu dan nilai diskrit. Algoritma yang merupakan perbaikan dari algoritma ID3 ini memperbaiki beberapa kelemahan pada algoritma tersebut, diantaranya (Lendra & Firdaus, 2020):

1. Mampu menangani atribut dengan tipe diskrit atau kontinu.
2. Mampu menangani atribut nilai yang hilang
3. Dapat memangkas cabang.

Algoritma C4.5 menggunakan tiga formula *entropy*, *gain*, dan *split info*, dan *gain ratio* dalam proses seleksi atribut-atribut data. Adapun bentuk persamaan-persamaan yang digunakan pada formula tersebut seperti terlihat pada persamaan (1) sampai persamaan (4) berikut (Dornubari & Angelawaale, 2021):

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

$$Gain(D, v) = Ent(D) - \sum_{i=1}^{n} \frac{|D_i|}{D} Ent(D_i) \quad (2)$$

$$SplitInfo(p, A) = -\sum_{i=1}^{n} \frac{|D_i|}{D} log_2 \frac{|D_i|}{D} \quad (3)$$

$$GainRatio = \frac{Gain(p,T)}{SplitInfo(p,T)} \quad (4)$$

2.6 Algoritma Random Forest

Algoritma *random forest* merupakan algoritma *machine learning* yang populer karena memiliki kinerja yang baik jika dibandingkan dengan algoritma klasifikasi lainnya, karena memiliki kelebihan dalam hal ketahanannya terhadap *overfitting*. *Random forest* merupakan algoritma yang termasuk ke dalam kategori *supervised learning* dan memiliki konsep pembelajaran *ensemble*, yang menggabungkan beberapa *classifier* untuk memecahkan masalah yang sangat kompleks dengan memperhatikan efisiensi dari model yang dihasilkan. *Classifier* pada *random forest* berisi pohon-pohon angka di dalam beberapa subset dataset dan menghitung rata-rata nilai tersebut untuk meningkatkan akurasi prediksi dari dataset tersebut (Bhardwaj & Chaurasia, 2022).

Yang membedakan antara *random forest* dan C4.5 adalah proses penambahan tahapan *random sub-setting* sebelum dilakukannya pembentukan setiap *tree*. Adapun langkah-langkah pembentukan *tree* dalam algoritma *random forest* adalah sebagai berikut (Azhari et al., 2021):

1. Tahapan *bootstrap*

   Tarik contoh acak dengan permulihan berukuran n dari gugus data training

2. Tahapan *random sub-setting*

   Susun tree berdasarkan data tersebut, namun pada setiap proses pemisahan pilih secara acak m < d peubah penjelas, dan lakukan pemisahan terbaik.

3. Ulangi langkah a-b sebanyak k kali sehingga diperoleh k buah tree acak

4. Lakukan pendugaan gabungan berdasarkan k buah *tree* tersebut (misal menggunakan *majority vote* untuk kasus klasifikasi, atau rata-rata untuk kasus regresi)