

PENERAPAN METODE DECISION TREE UNTUK PRAKIRAAN CUACA KOTA BEKASI

Kristopher Halim ¹⁾ Dyah Erny Herwindiati ²⁾ Tri Sutrisno ³⁾

¹⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S.Parman No. 1, Jakarta Barat 11440 Indonesia
email : Kristopher.825190019@stu.untar.ac.id ¹⁾ ,dyahh@fti.untar.ac.id ²⁾ ,tris@fti.untar.ac.id³⁾

ABSTRACT

To predict the weather requires a lot of weather data variables, many influencing factors such as the large amount of data. Because this makes prediction accuracy and speed less accurate. So with that made a way to make predictive model research using several data mining techniques. The data to be used in this study were obtained from timeanddate.com, timeanddate.com is a site that provides data and information about daily weather conditions. The data used ranges from December 2021 to August 2022. The research aims to obtain weather classification patterns using the data mining classification algorithm, namely the C4.5 algorithm. The results of testing the C4.5 algorithm using the php website and program get an accuracy of 82%.

Key words

Algorithm, Classification, Data Mining, C4.5

1. Pendahuluan

Proses prakiraan cuaca memerlukan banyak komponen data cuaca, jumlah data yang besar. Ini menyebabkan ketepatan dan kecepatan prakiraan kurang terpenuhi. Untuk mencoba menyelesaikan masalah yang ada, para peneliti mencoba melakukan beberapa menggunakan beberapa teknik data mining.

Pertumbuhan pesat dari banyaknya jumlah data cuaca dapat mendatangkan kondisi yang biasa disebut "Rich of Data but Poor of Information" karena data cuaca yang terkumpul tidak dapat digunakan untuk pengambilan keputusan dan akan menjadi kuburan data. Karena masalah ini munculah data mining yang dapat berguna untuk mengolah banyaknya data data tersebut dan dapat digunakan untuk memprakirakan cuaca yang akan datang. Maka dari itu penulis sangat ingin mencoba melakukan penelitian penerapan data mining dalam memprediksi cuaca dan ingin membuktikan seberapa akurat hasil prediksi dari program machine learning dan algoritma C4.5. Banyaknya parameter dalam menentukan suatu cuaca menyebabkan ketepatan dan kecepatan dalam memprediksi cuaca kurang terpenuhi. akibatnya perlu digunakannya metode data mining untuk

melakukan prediksi cuaca agar mendapatkan hasil yang nantinya diharapkan akan lebih baik.

Berdasarkan rumusan masalah yang ada pada penelitian ini akan digunakan beberapa atribut yang akan digunakan untuk melakukan penelitian dengan metode decision tree yang berjumlah 5 atribut dan nantinya juga akan ada atribut yang akan dijadikan sebagai target atau label atribut yang digunakan adalah sebagai berikut suhu rata-rata, tekanan udara, kecepatan angin, kelembapan, dan juga cuaca yang terjadi pada hari itu. Untuk menentukan pohon keputusan akan digunakan algoritma C4.5 untuk menentukan setiap node yang ada pada pohon keputusan yang akan terbentuk nanti.

2. Dasar Teori

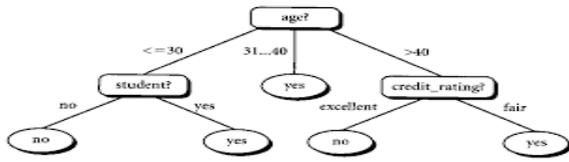
2.1 Pohon Keputusan

Pohon keputusan merupakan salah satu dari banyaknya metode machine learning, pohon keputusan juga suatu metode yang bersifat prediktif, yang pada intinya pohon keputusan akan mengubah suatu data yang tersedia menjadi sebuah keputusan, sebuah keputusan yang didapat ini didapatkan dari sebuah hubungan yang terdapat dalam atribut-atribut yang digunakan dalam sebuah dataset.

Pohon keputusan memiliki tiga node yang dapat dijabarkan sebagai berikut:

- *Root node*
Root node merupakan simpul yang terdapat dipaling atas dan akan dijadikan acuan dalam pengambilan keputusan.
- *internal node*
internal node merupakan simpul dalam pohon keputusan yang berguna untuk melakukan pilihan yang akan diambil, banyaknya cabang yang masuk maksimal hanya satu tetapi cabang yang dihasilkan dapat lebih dari Saturday
- *Leaf node*
Leaf node merupakan hasil dari pohon keputusan atau sebuah simpul yang berada dipaling bawah atau juga hasil dari pohon keputusan yang terbentuk.

Berikut ini adalah ilustrasi dari pohon keputusan.



Gambar 1 ilustrasi pohon keputusan

2.2 Algoritma C4.5

Algoritma C4.5 merupakan sebuah algoritma yang klasifikasi dalam melakukan proses data mining. Algoritma C4.5 memiliki cara kerja yang terdiri dari beberapa Langkah-langkah yang nantinya akan membentuk sebuah node. Berikut adalah Langkah-langkah untuk menggunakan algoritma C4.5:

- **Langkah Pertama**
Menentukan label dari salah satu atribut yang tersedia pada dataset. Contoh dataset pada Gambar 2.

suhu_rata_rata	Kecepatan_Angin_maksimum	tekanan_udara	humidity	uraian
Tinggi	Normal	normal	Normal	tidakhujan
Tinggi	Tinggi	normal	Normal	tidakhujan
Tinggi	Tinggi	tinggi	Lembap	tidakhujan
Tinggi	Normal	normal	Normal	mendung
Normal	Tinggi	tinggi	Lembap	hujan
Tinggi	Tinggi	tinggi	Lembap	tidakhujan
Tinggi	Tinggi	tinggi	Normal	tidakhujan
Tinggi	Tinggi	tinggi	Lembap	tidakhujan
Tinggi	Tinggi	tinggi	Lembap	tidakhujan
Normal	Normal	tinggi	Lembap	hujan
Tinggi	Normal	tinggi	Normal	tidakhujan
Tinggi	Tinggi	tinggi	Normal	tidakhujan

Gambar 2 Contoh Dataset

Pada contoh diatas atribut uraian akan dijadikan label atau target.

- **Langkah kedua**
Melakukan perhitungan dari data yang telah dijabarkan sebagai berikut Gambar 3.

no		jumlah kasus	hujan	tidak hujan	mendung	entropy	gain	
1	total	365	110	193	52	1.408097397		
	suhu rata						0.070491526	
			tinggi	100	11	75	14	1.058675005
			normal	265	110	118	37	1.442862803
	kecepatan angin							0.002755691
			normal	297	100	160	37	1.383868317
			tinggi	68	21	33	14	1.499129889
	tekanan udara							0.017905705
			tinggi	87	41	38	8	1.350059844
			normal	158	45	90	23	1.383261172
			rendah	120	35	65	20	1.428412468
	humidity							0.239658635
			lembap	216	116	76	24	1.364112565
			normal	149	5	117	27	0.884777412

Gambar 3 Perhitungan data

Tabel ini berfungsi untuk membantu melakukan penulisan hasil yang telah di dapat. Angka yang di masukkan di dalam tabel adalah data yang di dapatkan dari data set. Dengan memasukkan isi dari parameter class pada baris pertama sebagai indicator.

- **Langkah Ketiga**
Setelah pembuatan tabel, maka dilanjutkan dengan pencarian hasil Entropy, ini diperlukan untuk menghitung nilai gain. Berikut adalah rumus yang digunakan dalam mencari nilai Entropy:

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

Keterangan :

S : Banyaknya kasus

A : Atribut

N : Jumlah partisi dari banyaknya kasus

π : proporsi dari Si terhadap S

- **Langkah Keempat**
Setelah melakukan pencarian Entropy pada setiap atribut, di lanjutkan dengan pencarian gain. Dengan rumus sebagai berikut.

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

Keterangan:

S: Himpunan Kasus

A: Atribut

n: Jumlah partisi atribut A Jumlah kasus pada partisi ke i Jumlah kasus dalam S

π : proporsi dari Si terhadap S

Si : Jumlah kasus pada partisi sampai data ke-1

S : Jumlah kasus dalam S

2.3 Confusion Matrix

Pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.

2.4 Knowledge in Discovery Database

Knowledge in Discovery Database yang juga diketahui merupakan salah satu metode yang banyak dan sering dipakai untuk melakukan tahap tahapan dalam data mining. Ada beberapa Langkah dalam Knowledge in Discover Database seperti pemilihan data, preprocessing data, transformasi data, data mining, dan juga evaluasi sehingga akan dihasilkan pengetahuan atau output yang terbentuk dari metode Knowledge in Discover Database.

3. Hasil Percobaan

Dengan menggunakan metode Knowledge Discovery in Database. maka ada beberapa tahapan dan juga proses yang akan dilalui didalam penelitian ini untuk mendapatkan hasil atau output prediksi cuaca yang nantinya akan menjadi hasil dari penelitian yang telah dilakukan.

3.1 Pemilihan Data

Dataset yang dipakai pada penelitian ini berasal dari website timeanddate.com yang merupakan sebuah website untuk mendapatkan data historik cuaca yang nantinya akan dikumpulkan kedalam sebuah file excel.data yang digunakan pada penelitian ini dari bulan agustus 2021 – april 2023 yang berjumlah 638 data.dan lokasi yang akan digunakan adalah cuaca yang berasal dari kota Bekasi yang akan dipakai.

3.2 Preprocessing

Pada tahap data preprocessing nantinya akan dilakukan pembersihan dari data. seperti data yang tidak lengkap dan data yang isinya tidak tepat.setelah dilakukan proses preprocessing Sehingga didapatkan data sebanyak 638. data data yang telah didapatkan dapat dilihat pada tabel 1.

Tabel 1.Preprocessing

No	Suhu	Kecangin	Tekanan	Humidity	Cuaca
1	Tinggi	Normal	Normal	Normal	Hujan
2	Tinggi	Tinggi	Normal	Normal	Tidak hujan
3	Tinggi	Tinggi	Tinggi	Lembap	Hujan
...
638	Tinggi	Tinggi	Tinggi	Hujan	Hujan

3.3 Transformation

Tahap berikutnya peneliti mengubah dataset yang sudah kita miliki kedalam bentuk klasifikasi untuk nantinya memudahkan dalam pengolahan menggunakan algoritma yang digunakan.contohnya seperti suhu akan diklasifikasikan menjadi Tinggi,Normal sedangkan kecepatan angin akan diklasifikasikan menjadi Tinggi,Normal, sedangkan Tekanan udara juga diklasifikasikan menjadi tiga klasifikasi Tinggi,Normal,rendah.Humidity juga diklasifikasikan menjadi Normal,Lembap dan yang terakhir ada cuaca yang juga menjadi target atau label diklasifikasikan menjadi hujan dan tidak hujan.cuaca nantinya akan menjadi hasil akhir dari keputusan yang akan dibuat oleh pohon keputusan yaitu hujan atau tidak hujan. peneliti juga mencoba membagi data menjadi data training dan juga data testing dengan perbandingan 50:50 sehingga didapatkan data testing sebanyak 319 dan data train 319.Hasil Klasifikasi dapat dilihat pada Tabel 2.

Tabel 2. Hasil Klasifikasi

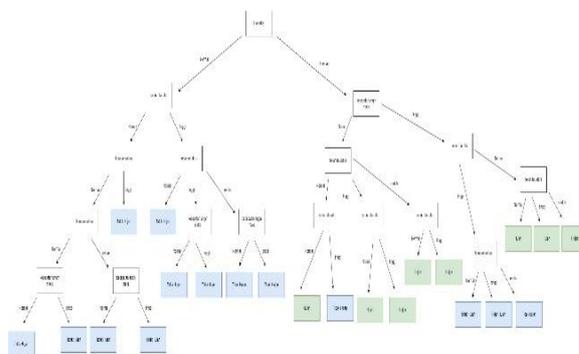
No	Atribut	Hasil Klasifikasi
1	Suhu	Tinggi,Normal
2	Kecepatan Angin	Tinggi,Normal
3	Tekanan Udara	Tinggi,Normal,Rendah

4	Humidity	Normal,Lembap
5	Cuaca	Hujan,Tidak Hujan

Hasil dari Klasifikasi berbagai atribut yang nantinya akan digunakan sebagai atribut pendukung untuk membuat pola pohon keputusan.

3.4 Hasil Pemodelan Decision Tree

Setelah dilakukan berbagai tahapan Knowledge Discovery in database didapatkan hasil dari model yang kemudian akan menghasilkan sebuah visualisasi pola pohon keputusan dengan data data yang sudah kita miliki dikarenakan algoritma C4.5 termasuk ke dalam algoritma decision tree. Pola dari visualisasi Pohon keputusan yang dihasilkan dapat dilihat pada gambar 4.



Gambar 4 Hasil Decision Tree

Berdasarkan gambar 3 didapatkan kesimpulan bahwa terbentuknya pohon keputusan dipengaruhi parameter humidity menjadi atribut yang paling penting dalam menentukan sebuah pola pohon keputusan diikuti dengan atribut lainnya seperti kecepatan udara,tekanan udara,dan suhu rata rata pada penelitian ini,dan juga terbentuklah rules yang berjumlah 22 yang berarti terdapat 22 hasil prediksi yang didapat dari pohon keputusan yang telah terbentuknya yang nantinya akan dijadikan sebuah acuan dalam melakukan hasil prediksi.

3.5 Evaluasi

Berdasarkan dari percobaan yang telah dilakukan dengan menggunakan program php yang sudah dibuat dan nantinya akan dilakukan percobaan sebanyak 5 kali yang bertujuan untuk melihat pembagian dataset mana yang nantinya akan mendapatkan hasil akurasi yang terbesar. dilakukan mulai dari pembagian membagi dataset menjadi 50:50,60:40,70:30,80:20,dan yang terakhir adalah 90:10. Pertama dengan membagi 50:50 didapatkan data latih 319 dan data test 319. maka didapatkan kesimpulan metode decision tree mendapatkan akurasi yang cukup besar yaitu 82.4%

sehingga dengan hasil ini maka decision tree dapat dijadikan salah satu metode untuk melakukan prediksi karena memiliki performa yang bisa dikatakan cukup baik.

Presentase
 Persentase Data Training 50%, Data Testing 50%

50 %

Hasil Performance
 Hasil Akurasi : 82.445%

Gambar 5 Hasil Akurasi 50:50

Berdasarkan dari percobaan ke 2 yang telah dilakukan dengan menggunakan program php yang sudah dibuat dan menggunakan perbandingan data 60:40 atau data latih yang berjumlah 382 dan data test yang berjumlah 256. maka didapatkan kesimpulan metode decision tree mendapatkan akurasi yang cukup besar yaitu 80.4% sehingga dengan hasil ini maka data dengan perbandingan 60:40 memiliki akurasi yang dapat dikatakan cukup tinggi tetapi masih berada dibawah data perbandingan 50:50. hal ini dapat dikarenakan karena semakin banyak data training yang dipakai data yang adapun dapat mempengaruhi keputusan meski data yang digunakan lebih banyak.

Presentase
 Persentase Data Training 60%, Data Testing 40%

60 %

Hasil Performance
 Hasil Akurasi : 80.469%

Gambar 6 Hasil akurasi 60:40

Berdasarkan dari percobaan ke 3 yang telah dilakukan dengan menggunakan program php yang sudah dibuat dan menggunakan perbandingan data 70:30 atau dapat dijabarkan dengan data latih yang berjumlah 446 dan data test yang berjumlah 192. maka didapatkan kesimpulan metode decision tree mendapatkan akurasi yang cukup besar yaitu 77.6% sehingga dengan hasil ini maka data dengan perbandingan 70:30 memiliki akurasi yang dapat dikatakan cukup tinggi tetapi masih berada dibawah data perbandingan 60:40 dan 50:50.

Presentase
 Persentase Data Training 70%, Data Testing 30%

70 %

Hasil Performance
 Hasil Akurasi : 77.604%

Gambar 7 Hasil Akurasi 70:30

Berdasarkan dari percobaan ke 4 yang telah dilakukan dengan menggunakan program php yang sudah dibuat dan menggunakan perbandingan data 80:20 atau data latih 510 dan data test 128. maka didapatkan kesimpulan metode decision tree mendapatkan akurasi yang cukup besar yaitu 78.1% sehingga dengan hasil ini maka data dengan perbandingan 80:20 memiliki akurasi yang dapat dikatakan cukup tinggi berada di atas dari data perbandingan 70:30 tetapi masih berada dibawah data perbandingan 60:40 dan 50:50.

Presentase
 Persentase Data Training 80%, Data Testing 20%

80 %

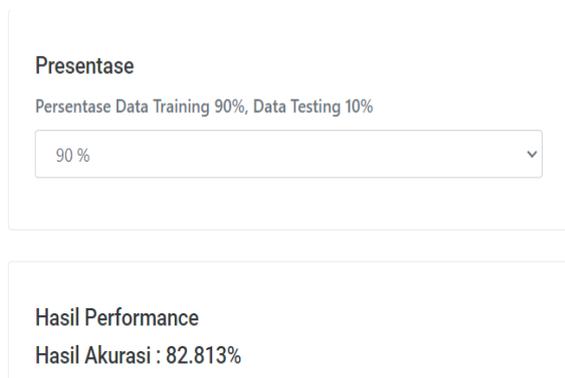
Hasil Performance
 Hasil Akurasi : 78.125%

Gambar 8 Hasil akurasi 80:20

Berdasarkan dari percobaan ke 5 yang telah dilakukan dengan menggunakan program php yang sudah dibuat dan menggunakan perbandingan data 90:10 atau data latih 574 dan data test 64. maka didapatkan kesimpulan metode decision tree mendapatkan akurasi yang cukup besar yaitu 82.8% sehingga dengan hasil ini maka data dengan perbandingan 90:10 memiliki akurasi yang dapat dikatakan cukup tinggi berada di atas akurasi dari empat percobaan sebelumnya.

Dengan adanya penelitian ini membuktikan bahwa sebuah proses data mining terutama yang digunakan pada penelitian ini yaitu menggunakan decision tree memerlukan jumlah data yang dapat dikatakan jumlahnya cukup banyak. banyaknya data ini dapat mempengaruhi sebuah hasil prediksi data proses data mining dan juga akurasi dari sebuah data mining. karena dengan banyaknya data, sebuah proses data mining akan lebih mudah lagi untuk menambah knowledge nya atau juga dengan membaca pattern dari banyaknya data yang sudah ada. yang nantinya akan dikeluarkan berupa output yang sesuai dan diinginkan oleh

peneliti.kekonsistenan data yang ada pada dataset juga diperlukan karena jika data yang ada tidak konsisten akan sulit bagi suatu proses data mining untuk memahaminya.



Gambar 9 Hasil Akurasi 90:10

Perbandingan data dengan pembagian 90:10 mendapatkan akurasi yang paling tinggi.

4. Kesimpulan

Berdasarkan dari hasil uji coba yang dilakukan diperoleh sebuah Kesimpulan yang didapatkan setelah melakukan lima kali percobaan dalam penelitian ini bahwa metode decision tree dapat dijadikan salah satu metode untuk melakukan prediksi atau prakiraan cuaca. karena mendapatkan hasil akurasi yang cukup baik serta berbagai macam atribut yang dapat digunakan dan juga banyaknya jumlah data sangat mempengaruhi hasil dari akurasi dikarenakan banyaknya data akan melatih model yang dibuat sehingga akan didapatkan hasil yang lebih baik lagi dalam penelitian ini. Didapatkan juga bahwa humidity menjadi salah satu faktor yang paling penting dalam penentuan hasil cuaca yang akan terjadi. dan juga program yang sudah dibuat berfungsi dengan baik dan sudah berjalan sebagaimana dengan fungsinya. dan dapat melakukan perhitungan akurasi. program juga dapat membaca dan juga membagi dataset. dan juga dengan rentang waktu lebih dari tahun decision tree dapat berjalan dan memiliki akurasi yang sangat baik.

REFERENSI

- [1] Subekti Mujiasih, "PEMANFAATAN DATA MINING UNTUK PRAKIRAAN CUACA"(Jakarta;BMKG,2011)
- [2] Yohanes Leonardus Dwi Pradipta, "ANALISIS FLUKTUASI HARGA SAHAM BLUE CHIPS DENGAN MENGGUNAKAN ALGORITMA C4.5"(Jakarta:tarumanagara,2022).
- [3] William Wijaksana," Pembuatan Aplikasi Berbasis Website Untuk Rekomendasi Fakultas dengan Algoritma C4.5"(Jakarta:tarumanagara,(Jakarta:tarumanagara,2022).
- [4] Adhika Novandy," PENERAPAN ALGORITMA KLASIFIKASI DATA MINING C4.5 PADA DATASET CUACA WILAYAH BEKASI"(Bekasi;BSI,2017)
- [5] Broto Poernomo T.P.dkk, PENERAPAN DATA MINING UNTUK PRAKIRAAN CUACA DI KOTA MALANG

- MENGGUNAKAN ALGORITMA ITERATIVE DICHOTOMISER TREE (ID3)",(Malang:STMIK,2017).
- [6] Syaiful Bahri," METODE KLASIFIKASI DECISION TREE UNTUK MEMREDIKSI JUARA ENGLISH PREMIER LEAGUE",(Medan, Universitas Pembangunan Panca Budi,2020).
- [7] Charles Yuliansen," Perancangan Aplikasi Prediksi Masa Studi Mahasiswa dengan Metode Naive Bayes dan C4.5"(Jakarta:tarumanagara,(Jakarta:tarumanagara,2022).
- [8] Simon, S. and Trisnawarman, D., 2014. Aplikasi Prediksi Status Registrasi Mahasiswa Baru Menggunakan Metode Naive Bayes dan Algoritma C4. 5. Jurnal Ilmu Komputer dan Sistem Informasi, 2(2), pp.216-219.
- [9] Mendrofa, Y. (2019). Implementasi Algoritma C4.5 Untuk Memprediksi Tingkat Kerusakan Akibat Banjir (Studi Kasus : BPBD Prov.Sumut). Jurnal Pelita Informatika, 7(4), 584-592.)

Kristopher Halim, seorang mahasiswa dari program studi sistem informasi fakultas teknologi informatika di Universitas Tarumanagara

Tri Sutrisno, Memperoleh gelar S.Si dari Universitas Diponegoro tahun 2011. Kemudian memperoleh gelar M.Sc.dari Universitas Gadjah Mada pada tahun 2015. Dan saat ini aktif sebagai Dosen Tetap Program Studi Teknik Informatika Universitas Tarumanagara Jakarta.

Dyah Erny Herwindiati, Memperoleh gelar Doktor MIPA,Matematika & Ilmu Pengetahuan Alam dari Institut Teknologi Bandung,Indonesia tahun 2006. saat ini sebagai Dekan Fakultas Teknologi Informasi Universitas Tarumanagara.