

---

## KLASIFIKASI POHON KEPUTUSAN DENGAN ALGORITMA C45 UNTUK KASUS PEMILIHAN PRODUK

Oleh

Deborah Kurniawati<sup>1</sup>, Endang Wahyuningsih<sup>2</sup>, Diva Putri Welfianti<sup>3</sup>

<sup>1,2,3</sup> Universitas Teknologi Digital Indonesia

Email: <sup>1</sup>[debbie@utdi.ac.id](mailto:debbie@utdi.ac.id), <sup>2</sup>[ayuningsih@utdi.ac.id](mailto:ayuningsih@utdi.ac.id)

---

### **Article History:**

Received: 09-09-2023

Revised: 17-09-2023

Accepted: 12-10-2023

### **Keywords:**

Hormones, *Azolla sp*,  
fermentation, growth,  
production

**Abstract:** *Customer service pada sebuah perusahaan menjadi garda utama bagi layanan perusahaan ke konsumen, terutama terkait dengan produk yang dihasilkan oleh perusahaan. Setiap konsumen menginginkan produk yang tepat bagi dirinya, apalagi terkait dengan produk kecantikan. Banyaknya jenis dan varian produk, beragamnya masalah konsumen, dan terbatasnya jumlah pegawai customer service membuat waktu yang dibutuhkan untuk konsultasi produk menjadi lebih banyak. Untuk itu dibutuhkan model yang dapat memberikan pilihan produk yang tepat bagi konsumen. Model klasifikasi pemilihan produk yang tepat bagi konsumen menjadi salah satu solusi.*

*Model klasifikasi yang dibuat menggunakan algoritma C45 untuk membangun pohon keputusan yang akan digunakan dalam klasifikasi pemilihan produk. Kriteria yang digunakan dalam klasifikasi ini adalah jenis kulit, masalah kulit utama, masalah kulit wajah lainnya, dan umur. Rekomendasi yang diberikan oleh sistem berupa jenis masker yang terdiri dari 3 varian. Data yang digunakan untuk pembuatan pohon keputusan berjumlah 206 dan pohon keputusan yang dihasilkan bersifat statis.*

*Dari data yang ada dihasilkan 114 rule keputusan dengan akurasi hasil sebesar 96%.*

---

## PENDAHULUAN

*Customer service pada sebuah perusahaan menjadi garda terdepan bagi layanan perusahaan ke konsumen, terutama terkait dengan produk yang dihasilkan oleh perusahaan. Setiap konsumen menginginkan produk yang tepat bagi dirinya, apalagi terkait dengan produk kecantikan. Selama ini perusahaan belum memiliki fasilitas konsultasi secara langsung dengan dokter melainkan hanya tanya jawab mengenai produk-produk yang cocok melalui customer service. Banyaknya jenis dan varian produk, dan beragamnya masalah konsumen membuat waktu yang dibutuhkan untuk konsultasi produk menjadi lebih banyak. Selain itu, antrian konsultasi juga menjadi hal yang tidak dapat dihindari karena jumlah customer service yang terbatas. Keterbatasan customer service ditambah dengan tugas customer service yang bukan hanya untuk konsultasi produk, tetapi juga menangani masalah stok produk.*

Algoritma C45 merupakan salah satu algoritma yang dapat digunakan dalam data mining sebagai pengklasifikasi pohon keputusan yang dapat digunakan untuk menghasilkan keputusan berdasarkan sampel data tertentu (1). Pohon keputusan dapat diimplementasikan dalam system otomatis untuk membantu dalam pengambilan keputusan. Misalnya dalam system pendukung keputusan, pohonkeputusan dapat memberikan saran atau rekomendasi berdasarkan input dari pengguna atau data yang dimiliki. Algoritma C45 merupakan pengembangan dari ID3 (2). Pohon keputusan dengan algoritma C45 dapat diterapkan untuk menganalisa prediksi (3), penentuan pola (4), penilaian (5), klasifikasi (6,7), dan lainnya.

### LANDASAN TEORI

Algoritma C4.5 adalah suatu deretan Algoritma untuk permasalahan klasifikasi didalam sebuah mesin dan himpunan data. Dengan nilai data yang bervariasi, dimana kejadian diuraikan oleh koleksi atribut dan mempunyai salah satu dari satu set kelas yang eksklusif. Algoritma C4.5 merupakan generasi baru dari Algoritma ID3 yang dikembangkan oleh J. Ross Quinlan pada tahun 1983. Untuk membuat sebuah pohon keputusan, Algoritma ini dimulai dengan memasukan training sample ke dalam simpul akar pada pohon keputusan. *Training sample* adalah sampel yang digunakan untuk membangun model *classifier* dalam hal ini pohon keputusan kemudian sebuah atribut dipilih untuk mempartisi sampel ini. Untuk tiap nilai yang dimiliki atribut ini, sebuah cabang dibentuk. Setelah cabang terbentuk maka subset dari himpunan data yang atributnya memiliki nilai yang bersesuaian dengan cabang tersebut dimasukan ke dalam simpul yang baru.

Secara umum Algoritma C45 memiliki tahapan utama yaitu:

#### 1. Pilih atribut sebagai akar

Memilih atribut sebagai akar bisa dihitung dengan melihat nilai *gain* dari masing-masing atribut. Untuk mendapatkan nilai *gain* perlu dilakukan mendapatkan nilai *entropy* dengan persamaan 1.

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

Keterangan persamaan (1):

S = Himpunan Kasus

k = banyaknya partisi S

P<sub>i</sub> = hasil bagi antara jumlah pada setiap nilai dengan n

Setelah mendapatkan nilai *entropy* selanjutnya mencari nilai *gain*, nilai *gain* tertinggi nantinya akan menjadi akar yang pertama. Bentuk persamaan untuk mendapatkan nilai *gain* bisa dilihat pada persamaan 2.

$$\text{Gain (S, A)} = \text{entropy (S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} \log_2(s_i) \quad (2)$$

Keterangan persamaan (2) :

S = ruang sampel data

A = atribut

S<sub>i</sub> = jumlah sampel pada partisi ke-i

2. Ulangi perhitungan hingga semua data terbagi.

Proses pengulangan pada metode *pohon keputusan* ini akan terhenti jika:

- a. Semua data telah terbagi rata
- b. Tidak ada lagi atribut yang bisa dibagi lagi
- c. Tidak ada data *record* dalam cabang kosong.

Pohon keputusan adalah struktur pohon seperti diagram alur, di mana setiap simpul internal (node non daun) menunjukkan pengujian pada atribut, setiap cabang mewakili hasil pengujian, dan setiap simpul daun (atau simpul terminal) memegang label kelas (8). Simpul paling atas dalam sebuah pohon adalah simpul akar. Pohon keputusan dapat digunakan untuk klasifikasi. Diberikan sebuah tupel,  $X$ , dimana label kelas terkait tidak diketahui, nilai atribut dari tupel diuji terhadap pohon keputusan. Jalur ditelusuri dari akar ke simpul daun, yang menyimpan prediksi kelas untuk tupel itu. Pohon keputusan dapat dengan mudah diubah menjadi aturan klasifikasi.

## METODE PENELITIAN

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut,

1. Analisa permasalahan

Analisa permasalahan difokuskan pada kriteria yang dipilih untuk pengklasifikasian produk, sesuai dengan kegunaan dan bahan-bahan yang terdapat pada produk. Untuk pemilihan produk digunakan 4 kriteria yaitu jenis kulit, permasalahan utama kulit, permasalahan kulit lainnya, dan usia konsumen.

2. Pengumpulan data

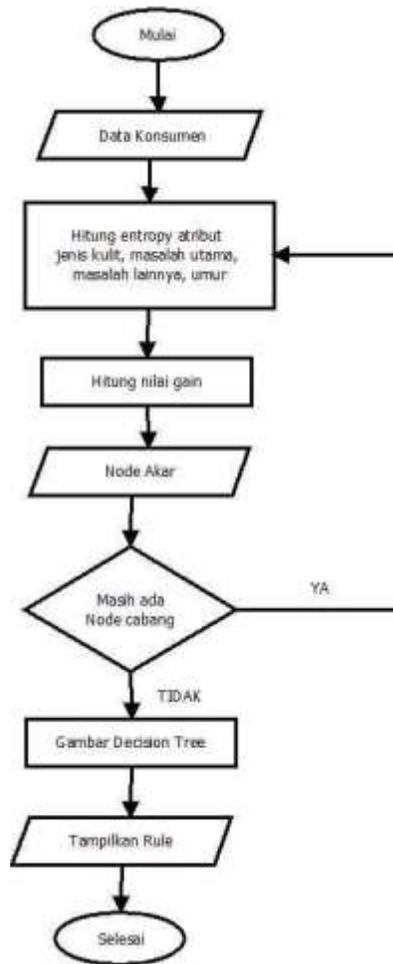
Data yang dikumpulkan terdiri atas data konsumen terkait dengan atribut yang digunakan pada model. Atribut yang digunakan terdiri dari data jenis kulit, masalah utama kulit, masalah kulit lainnya, dan umur.

3. Perancangan model

Pada tahap perancangan model akan dilakukan dengan membentuk pohon keputusan yang mengimplementasikan algoritma C45. Tahap yang dilakukan pada Langkah ini dapat dilihat pada Gambar 1. Pembuatan model menggunakan data konsumen, dengan menghitung nilai entropy dan nilai gain. Nilai *gain* tertinggi akan dijadikan node akar untuk membangun pohon keputusan. Selanjutnya mencari node cabang. Jika node cabang masih ada maka diulangi dengan perhitungan Algoritma C45 kembali sampai tidak ada lagi node cabang yang dapat terbentuk. Setelah pohon keputusan terbentuk maka diperoleh *rule* untuk rekomendasi produk.

4. Pengujian model

Pengujian model dilakukan dengan membandingkan hasil rekomendasi produk yang diberikan *customer service* sebelum menggunakan model dan hasil rekomendasi yang diperoleh dari model yang dibangun.



**Gambar 1. Alur perancangan model**

**HASIL DAN PEMBAHASAN**  
**Implementasi Algoritma C45**

Algoritma C45 digunakan untuk membuat pohon keputusan dalam penentuan produk bagi konsumen. Pilihan produk yang disediakan ada 4 buah yaitu produk K, produk M, produk C, dan produk H. Dengan menggunakan 206 baris data sebagai data latih, diperoleh fakta seperti yang tampak di Tabel 1.

**Tabel 1. Data latih**

Atribut		Jumlah	Produk K	Produk M	Produk C	Produk H
		206	70	90	33	13
Jenis kulit	normal	79	0	79	0	0
	Kering	28	1	11	3	13
	Berminyak	69	69	0	0	0
	Sensitive	30	0	0	30	0
Masalah utama						
		28	3	25	0	0
		42	14	16	8	4

	Bekas jerawat	54	22	17	8	7
	Pori-pori	31	8	23	0	0
	Jerawat	37	21	2	14	0
	Flek hitam	12	2	7	3	0
	keriput	2	0	0	0	2
Masalah lainnya	Tidak ada	44	7	35	2	0
	Beruntusan	37	12	16	7	2
	Bekas jerawat	47	17	14	12	4
	Pori-pori	43	20	12	7	4
	Jerawat	21	8	6	5	2
	Flek hitam	11	5	5	0	1
	keriput	3	1	2	0	0
umur	12-25 tahun	206	70	90	33	13
	26-50 tahun	0	0	0	0	0
	>51 tahun	0	0	0	0	0

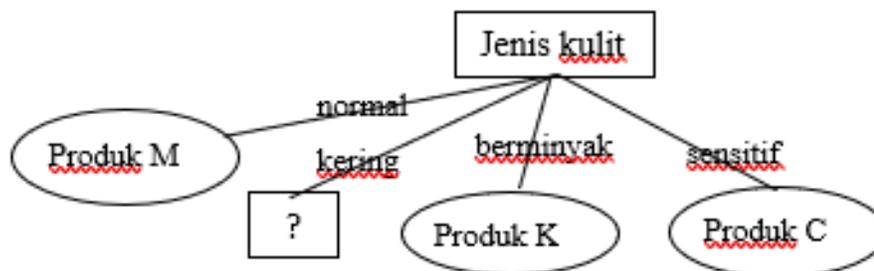
Dengan menggunakan persamaan 1 dan 2 maka nilai entropy dan gain untuk Node 1 yang diperoleh dapat dilihat pada Tabel 2.

**Tabel 2. Nilai entropy dan gain node 1**

Atribut	Jumlah	Produk K	Produk M	Produk C	Produk H	Nilai Entropy	Nilai Gain
	206	70	90	33	13	1,725882507	
<b>Jenis Kulit</b>							1,51378743
Normal	79	0	79	0	0	0	
Kering	28	1	11	3	13	1,560413792	
Berminyak	69	69	0	0	0	0	
Sensitif	30	0	0	30	0	0	
<b>Masalah utama (masalah1)</b>							0,76434133
Tidak ada	28	3	25	0	0	0,491237342	
Beruntusan	42	14	16	8	4	1,83748483	
Bekas jerawat	54	22	17	8	7	1,04774499	
Pori-pori	31	8	23	0	0	0,823811633	
Jerawat	37	21	2	14	0	0,22753802	
Flek hitam	12	2	7	3	0	1,384431504	
Keriput	2	0	0	0	2	0	
<b>Masalah lainnya (masalah2)</b>							0,36306756
Tidak ada	44	7	35	2	0	0	
Beruntusan	37	12	16	7	2	1,73185784	
Bekas jerawat	47	17	14	12	4	1,856516783	
Pori-pori	43	20	12	7	4	1,772553211	

Jerawat	21	8	6	5	2	1,862821448	
Flek hitam	11	5	5	0	1	1,028265907	
Keriput	3	1	2	0	0	0,918295834	
<b>Umur</b>							0
12-25 tahun	206	70	90	33	13	1,725882507	
26-50 tahun	0	0	0	0	0	0	
>51 tahun	0	0	0	0	0	0	

Dengan menggunakan beberapa istilah yang ada di struktur data *tree* (pohon) yaitu *root*, level dan *node*, maka sesuai dengan perhitungan yang ada pada Tabel 2 yang menjadi *root* pada pohon keputusan yang dibentuk adalah Jenis kulit. Hal ini dikarenakan jenis kulit memiliki nilai gain tertinggi yaitu 1,51378743. Untuk menentukan node-node pada level berikutnya digunakan nilai atribut dari jenis kulit, yaitu Normal, Kering, Berminyak, dan Sensitif. Dari ke empat nilai tersebut diperoleh nilai entropy tertinggi yaitu 1,560413792 pada nilai atribut kering. Hal ini menyebabkan nilai atribut kering akan menjadi *root* bagi level berikutnya. Nilai entropy dari tiga nilai atribut lainnya bernilai 0 sehingga untuk jenis kulit tersebut produk yang dipilih sudah dianggap sesuai. Hal ini akan menyebabkan pada pohon keputusan untuk nilai atribut tersebut tidak memerlukan level berikutnya, artinya pilihan produk yang ada sudah menjadi alternatif terpilih, atau dengan kata lain produk yang ada menjadi *node* akhir (*leaf*) pada pohon keputusan. Dari hal tersebut maka pohon keputusan awal yang dapat dibangun dapat digambarkan seperti yang ada pada Gambar 2.



**Gambar 2. Pohon keputusan node akar**

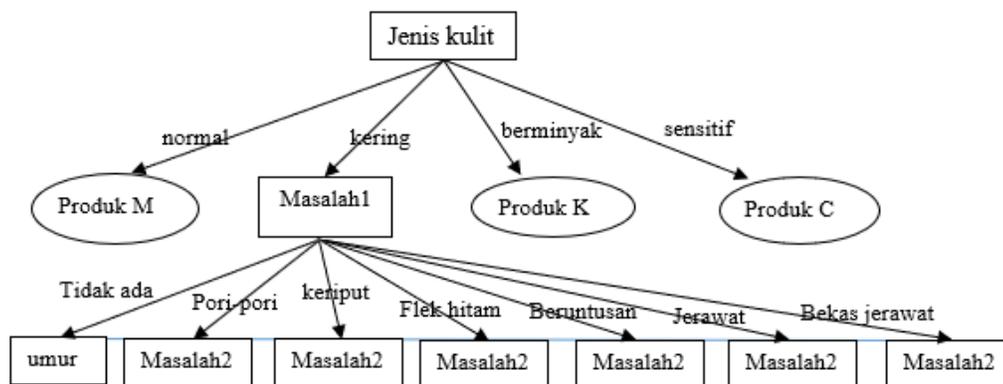
Langkah berikutnya adalah menentukan node induk yang akan diletakkan pada jalur kering. Dengan menggunakan persamaan 1 dan 2 maka nilai entropy dan gain untuk node di jalur kering dapat dilihat pada Tabel 3.

**Tabel 3. Nilai entropy dan gain jalur kering**

Atribut	Jumlah	Produk K	Produk M	Produk C	Produk H	Nilai Entropy	Nilai Gain
	28	1	11	2	13	1,560413792	
<b>Masalah1</b>							1,431498061
Tidak ada	6	0	6	0	0	0	
Beruntusan	5	0	1	0	4	0,721928095	
Bekas jerawat	7	0	0	0	7	0	
Pori-pori	4	0	4	0	0	0	
Jerawat	1	1	0	0	0	0	

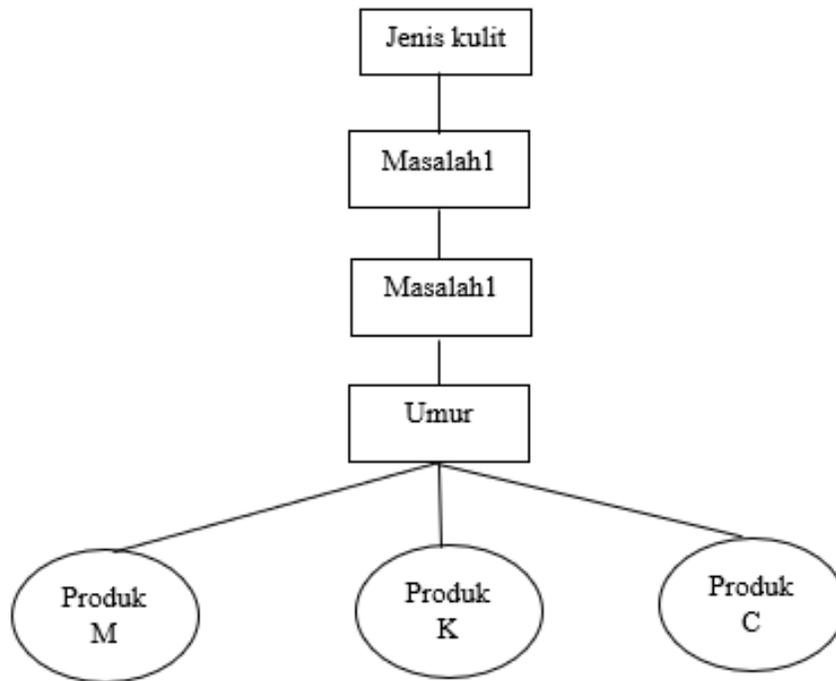
Flek hitam	3	0	0	3	0	0	
Keriput	2	0	0	0	2	0	
<b>Masalah2</b>							0,495546596
Tidak ada	6	0	5	1	0	0,650022422	
Beruntusan	4	0	1	1	2	1,5	
Bekas jerawat	7	1	2	0	4	1,378783493	
Pori-pori	6	0	1	1	4	1,251629167	
Jerawat	2	0	0	0	2	0	
Flek hitam	3	0	2	0	1	0,918295835	
Keriput	0	0	0	0	0	0	
<b>Umur</b>							0
12-25 tahun	28	1	11	3	13	1,560413792	
26-50 tahun	0	0	0	0	0	0	
>51 tahun	0	0	0	0	0	0	

Dari Tabel 3 terlihat nilai gain terbesar secara berurutan adalah atribut Masalah1, yaitu sebesar 1,431498061, Masalah2, dan atribut Umur. Hal ini akan membentuk jalur baru dengan membuat node yang berada pada level 1 di jalur adalah Masalah1, dan node yang berada di level 2 adalah Masalah2, dan di level terakhir adalah Umur. Dengan demikian, pohon keputusan sampai level 2 yang dapat dibentuk dapat dilihat pada Gambar 3.



**Gambar 3. Pohon keputusan sampai level 2**

Pada Gambar 3 terlihat jalur jenis kulit- kering-masalah1-tidakada dapat langsung menuju atribut Umur. Untuk jalur yang lain akan diteruskan ke atribut berikutnya yaitu Masalah2. Penentuan level berikutnya akan dilakukan dengan cara yang sama pada saat menentukan level 2. Level pohon keputusan yang akan terbentuk dapat digambarkan seperti yang tampak pada Gambar 4. Jalur yang akan terbentuk pada pohon keputusan sesuai dengan atribut dan nilai atribut yang digunakan.



**Gambar 4. Gambaran leveling pohon keputusan**

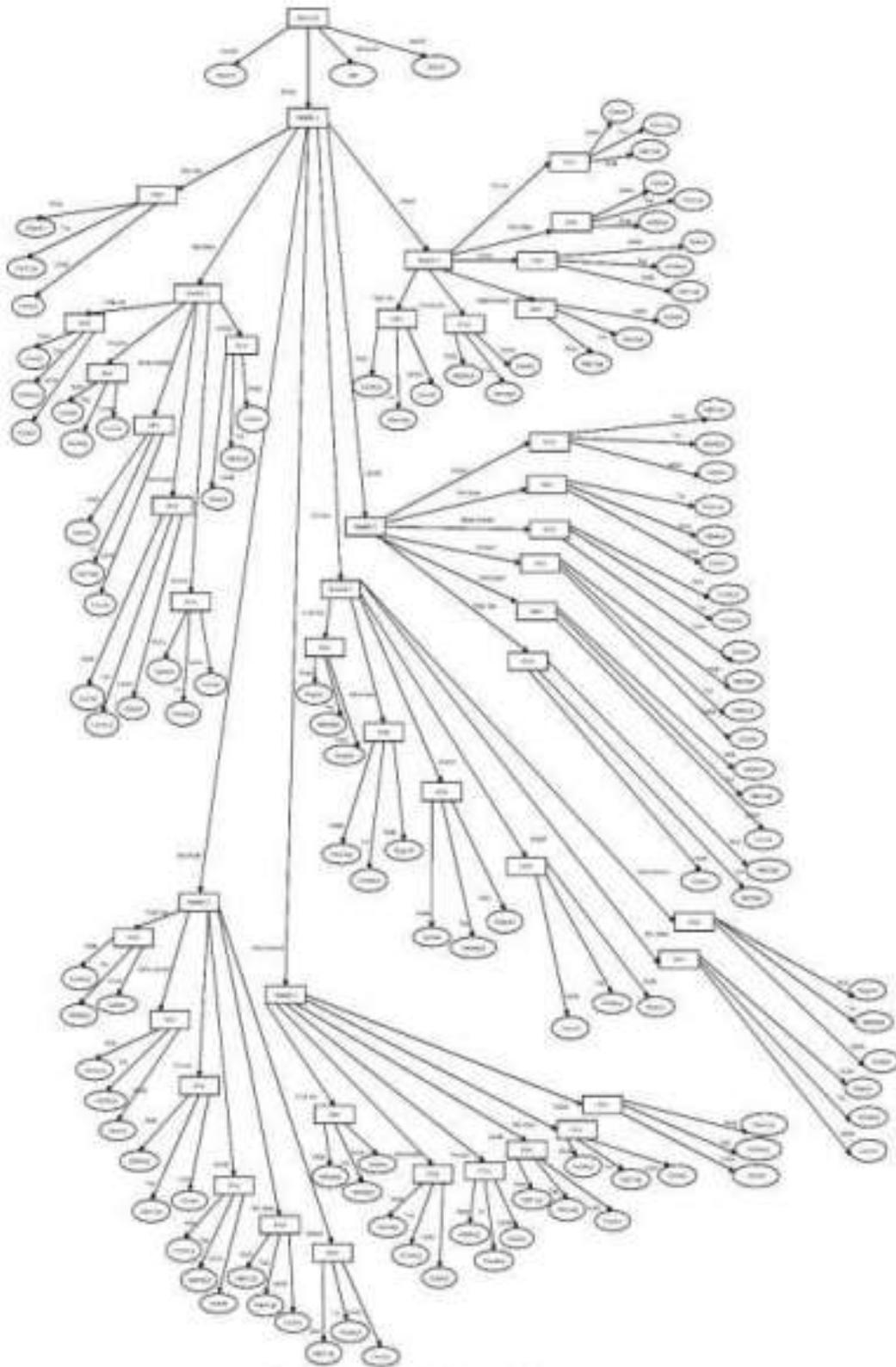
Pohon keputusan yang terbentuk akan menghasilkan rule berdasarkan jalur yang ada. Dari pohon keputusan yang terbentuk (Gambar 5) diperoleh 114 rule untuk menentukan rekomendasi produk. Dari 114 rule yang terbentuk, hanya 3 rule yang tidak terkait dengan jenis kulit kering. Hal ini sesuai dengan jalur yang terbentuk di Gambar 2. Rule-rule tersebut adalah

1. If jenis kulit normal THEN ProdukM
2. If jenis kulit berminyak THEN ProdukK
3. If jenis kulit Sensitif THEN ProdukC

Dari gambar 5 terlihat hampir seluruh jalur-jalur yang terbentuk berasal dari atribut jenis kulit-kering.

### **Pengujian Akurasi**

Pengujian hasil rekomendasi dilakukan dengan membandingkan rekomendasi yang langsung diberikan oleh *customer service* dengan hasil rekomendasi yang disarankan berdasarkan rule yang dihasilkan dari pohon keputusan. Untuk pengujian digunakan 29 data. Setelah dilakukan pengujian, diperoleh akurasi 96%, artinya dari 29 data hanya 1 data yang tidak sesuai.



Gambar 5. Pohon keputusan yang terbentuk

**KESIMPULAN**

Beberapa kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut,

1. Dengan menggunakan 4 atribut dari 206 data konsumen diperoleh 114 rule yang dapat digunakan untuk merekomendasikan produk yang dianggap sesuai dengan kondisi pelanggan.
2. Dari hasil pengujian diperoleh akurasi sebesar 96%.

**DAFTAR PUSTAKA**

- [1] Sumit Saha, 2018, What is the C4.5 algorithm and how does it work?, <https://towardsdatascience.com>
- [2] Eka Pandu Cynthia, Edi Ismanto, 2018, Metode Decision Tree Algoritma C.45 dalam mengklasifikasi data penjualan bisnis gerai makanan cepat saji, Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK), Volume 3.
- [3] Fitria Marisa, Anastasia L Maukar, 2022, Analisa Prediksi Varietas Buah Salak yang Sesuai dengan Lahan Daerah Kabupaten Banjarnegara menggunakan Algoritma C54, Jurnal Teknologi dan Manajemen Informatika, Vol 8 No 1.
- [4] Happy T., Dinda Pingsimoni, Indra Samsie, Nurul Aini, 2022, Penerapan Data Mining menggunakan Algoritma C45 dalam Menentukan Pola Pembelian Pakaian Syar'I yang Terjual, Jurnal Dipanegara Komputer Teknik Informatika (DIPAKOMTI), Vol 15 No 1.
- [5] Harry Dhika, Fitriana Destiawati, 2017, Penerapan Algoritma C45 untuk Penilaian Karyawan pada Restoran Cepat Saji, Proceeding Seminar Nasional Informatika dan Aplikasinya, Vol 3.
- [6] Nurlindasari Tamsir, Asrul Syam, Nirwana, Asmah Akhriana, Indo Intan, Vivi Rosida, 2021, Algorithm C4.5 in Classifying Health of Cat, Journal of Information Technology and Its Utilization, Volume 4, Issue 2.
- [7] Prameswari Reska Agami, Intan Yuniar Purbasari, Basuki Rahmat. 2021. Penentuan Penggunaan Lulur dan Masker Organik Sesuai dengan Diagnosa Jenis Kulit Wajah Menggunakan Metode Decision Tree Algoritma C4.5., tesis, UPN Veteran Jatim.
- [8] Han, J., Kamber, M., Pei, J., 2012, Data mining: concepts and techniques – 3rd ed, Elsevier Inc.