

---

**Analisis Faktor Yang Mempengaruhi Kepuasan Pegawai Dinas Pangan:  
Pendekatan Menggunakan Algoritma C4.5**

**Tongku Hamonangan Harahap<sup>1</sup>, Ibnu Rasyid Munthe<sup>2</sup>,  
Angga Putra Juledi<sup>3</sup>**

Sistem Informas, Fakultas Sains Dan Teknologi, Universitas Labuhanbatu<sup>1,2,3</sup>

Email : [tongkuhrp08@gmail.com](mailto:tongkuhrp08@gmail.com)<sup>1</sup>, [ibnurasyidmunthe@gmail.com](mailto:ibnurasyidmunthe@gmail.com)<sup>2</sup>,  
[anggaj19@gmail.com](mailto:anggaj19@gmail.com)<sup>3</sup>

Corresponding Author : [tongkuhrp08@gmail.com](mailto:tongkuhrp08@gmail.com)

**Abstract**

*The level of satisfaction is an important measure in evaluating the extent to which the needs and expectations of a person or group are met by a product, service, or experience. The concept is often used in a business context to measure how well a product or service meets customer expectations. The level of satisfaction can be measured through various methods such as surveys, interviews or analysis of consumer behavior data. The results of this satisfaction level evaluation provide valuable insights for companies in improving the quality of their products or services, as well as maintaining customer loyalty. Therefore, the author will conduct a study on the level of employee satisfaction Department of food using machine learning approach with C4.5 method. This study aims to explore the patterns and factors that significantly affect the level of employee satisfaction in the context of the Department of food. The C4.5 method was chosen because of its ability to handle complex and diverse data, as well as being able to provide insight into the relationship of complex and non-linear variables.*

**Keywords:** *Employee Satisfaction, Data Mining, Algorithm C4.5.*

**I. Pendahuluan**

Tingkat kepuasan merupakan ukuran yang penting dalam mengevaluasi sejauh mana kebutuhan dan harapan seseorang atau kelompok terpenuhi oleh suatu produk, layanan, atau pengalaman. Konsep ini sering kali digunakan dalam konteks bisnis untuk mengukur seberapa baik produk atau layanan memenuhi ekspektasi pelanggan. Tingkat kepuasan dapat diukur melalui berbagai metode seperti survei, wawancara, atau analisis data perilaku konsumen. Hasil dari evaluasi

tingkat kepuasan ini memberikan pandangan yang berharga bagi perusahaan dalam meningkatkan kualitas produk atau layanan mereka, serta mempertahankan loyalitas pelanggan. Selain itu, tingkat kepuasan juga memiliki implikasi yang luas dalam aspek sosial dan psikologis. Ketika seseorang merasa puas dengan pengalaman atau interaksi tertentu, hal ini dapat meningkatkan kesejahteraan psikologisnya dan membangun rasa percaya diri dalam mengambil keputusan. Masalah tingkat kepuasan

Pegawai sering kali menjadi perhatian utama bagi organisasi. Ketidakpuasan Pegawai dapat mempengaruhi produktivitas, loyalitas, dan retensi tenaga kerja dalam jangka panjang. Beberapa faktor yang dapat menyebabkan tingkat kepuasan rendah di antaranya adalah kurangnya komunikasi yang efektif antara manajemen dan Pegawai, ketidakjelasan mengenai peran dan tanggung jawab, serta kurangnya pengakuan terhadap kontribusi Pegawai. Ketika Pegawai merasa tidak dihargai atau tidak memiliki ruang untuk berkembang, hal ini dapat menyebabkan frustrasi dan mengurangi motivasi untuk berkepuasan dengan baik. Selain itu, masalah kepuasan Pegawai juga dapat muncul akibat lingkungan kerja yang tidak sehat atau kurang mendukung. Penelitian ini bertujuan untuk menggali pola dan faktor-faktor yang secara signifikan mempengaruhi tingkat kepuasan Pegawai dalam konteks dinas pangan.

## II. Landasan Teori Kepuasan Pegawai

Kepuasan bekerja merupakan suatu bentuk sikap dari rasa puas dan bahagia akan pekerjaan saat ini. Seseorang akan merasa puas di suatu perusahaan atau Lembaga yang dimana di dalamnya ada sebuah dedikasi dan di berikan kepercayaan oleh perusahaan atau Lembaga dalam melaksanakan tugas sesuai dengan keahliannya yang didalamnya terdapat rasa suka dan tanggung jawab untuk berusaha memberikan pelayanan dan perilaku yang terbaik untuk menunjukkan Kepuasan kerja yang baik.

## *Machine Learning*

Machine learning adalah cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan pengalaman tanpa perlu diprogram secara eksplisit. Teknologi ini berfokus pada pengembangan algoritma dan model matematis yang memungkinkan komputer untuk membuat keputusan atau melakukan prediksi berdasarkan pola yang ditemukan dalam data. Tujuan utama dari machine learning adalah untuk memungkinkan komputer untuk belajar secara otomatis dari pengalaman masa lalu dan mengadaptasi diri terhadap situasi baru tanpa perlu intervensi manusia yang terus-menerus.

## Algoritma C4.5

Metode C4.5 merupakan salah satu teknik populer dalam machine learning yang digunakan untuk masalah klasifikasi dan regresi. Teknik ini mengandalkan pembentukan sejumlah besar pohon keputusan (decision trees) yang bekerja secara kolektif untuk menghasilkan prediksi akhir. Setiap pohon keputusan dalam C4.5 dibangun secara acak dengan memilih sub-sampel acak dari data pelatihan dan fitur-fitur acak dari dataset.

Rumus dasar entropy

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

Keterangan: S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Setelah mendapatkan nilai entropy kemudian menghitung nilai gain dengan metode.

Rumus dasar gain

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

Keterangan :

S = Himpunan Kasus.

A = Atribut.

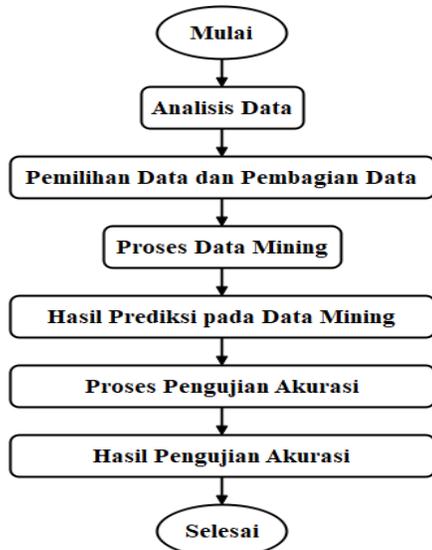
n = Jumlah Partisi Atribut.

$A | S_i |$  = Jumlah Kasus pada partisi-*i*.

$| S |$  = JumlahKasudalam S.

- Ulangi langkah ke-2 hingga semua tupel terpartisi.
- Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama, tidak ada atribut di dalam tupel yang dipartisi lagi dan tidak ada tupel di dalam cabang yang kosong.

### III. Metode Penelitian



Gambar 1. Kerangka Kerja Penelitian

### Analisis Data

Analisis data dalam konteks data mining merupakan proses penyelidikan dan penafsiran pola, hubungan, dan informasi yang terkandung dalam dataset yang besar dan kompleks. Tujuan utamanya adalah mengungkapkan pengetahuan yang berguna dan pemahaman mendalam tentang pola tersembunyi atau tren yang dapat digunakan untuk pengambilan keputusan. Proses analisis data melibatkan langkah-langkah seperti pemrosesan data, pemilihan atribut yang relevan, identifikasi pola, dan pembuatan model prediktif.

### Proses Data Mining

Proses data mining adalah serangkaian langkah sistematis yang dilakukan untuk menggali pengetahuan yang berharga dari dataset yang kompleks. Tahap pertama dalam proses ini adalah pemahaman bisnis dan karakteristik data yang relevan diidentifikasi. Langkah selanjutnya melibatkan pemilihan dan pemrosesan data, termasuk pembersihan data, penanganan nilai yang hilang, dan transformasi atribut. Setelah itu, tahap pemodelan melibatkan pemilihan metode analisis yang sesuai, seperti klasifikasi, regresi, clustering, atau asosiasi, dan pengembangan model berdasarkan dataset yang telah diproses sebelumnya. Tahap evaluasi dan validasi dilakukan untuk menilai Kepuasan model terhadap data yang tidak terlihat sebelumnya, dan jika hasilnya memuaskan, model tersebut dapat diimplementasikan untuk pengambilan keputusan atau prediksi. Terakhir, tahap implementasi dan

pemeliharaan melibatkan integrasi model ke dalam sistem yang ada dan pemantauan Kepuasannya seiring waktu. Proses ini memberikan kerangka kerja yang sistematis dan efektif untuk mendapatkan wawasan berharga dari data dan mendukung pengambilan keputusan yang informasional.

#### **Hasil Prediksi Pada Data Mining**

Hasil prediksi pada data mining merupakan output dari model analisis yang telah dikembangkan untuk memahami pola dan tren dalam dataset. Hasil ini memberikan gambaran tentang kemungkinan kejadian atau nilai di masa depan berdasarkan pembelajaran dari data historis. Dalam konteks klasifikasi, hasil prediksi dapat berupa penentuan kategori atau label tertentu untuk suatu observasi. Di regresi, hasilnya dapat berupa nilai prediksi berkelanjutan. Interpretasi hasil prediksi memegang peranan kunci dalam pengambilan keputusan, membantu organisasi untuk merencanakan strategi berbasis data, mengidentifikasi peluang atau risiko, dan meningkatkan efisiensi operasional.

#### **Hasil Pengujian Akurasi**

Hasil pengujian akurasi pada data mining menyediakan gambaran kritis tentang Kepuasan model analisis terhadap data yang tidak terlibat dalam proses pelatihan. Metode ini memungkinkan evaluasi objektif terhadap sejauh mana model dapat menggeneralisasi dan memberikan prediksi yang akurat pada situasi dunia nyata. Hasil pengujian akurasi diukur melalui berbagai metrik evaluasi, seperti akurasi keseluruhan, presisi, recall, F1-score, dan area di bawah

kurva ROC, yang memberikan wawasan mendalam tentang kemampuan model dalam mengklasifikasikan atau memprediksi data. Interpretasi hasil pengujian membantu pengambil keputusan dan analis data untuk menilai kehandalan model, mengidentifikasi potensi perbaikan, dan membuat keputusan strategis. Dengan memahami dan mengoptimalkan hasil pengujian akurasi, organisasi dapat memastikan bahwa model data mining yang dikembangkan dapat memberikan informasi yang dapat diandalkan untuk mendukung proses pengambilan keputusan yang lebih baik

#### **IV. Hasil dan Pembahasan Analisis Sistem**

Sistem ini dimulai dengan pengumpulan data yang relevan pada Dinas Pangan memerlukan struktur yang efisien untuk pengolahan data yang sesuai, dalam hal ini algoritma C4.5, berdasarkan pemahaman mendalam tentang karakteristik data dan potensi hasil yang akurat. Selanjutnya penelitian memerlukan data yang akurat sehingga penelitian ini dapat berjalan dengan baik dan mencapai nilai tujuan yang sesungguhnya dengan mengaplikasikan metode algoritma C4.5 untuk mengetahui faktor yang mempengaruhi kepuasan pegawai Dinas Pangan. Sehingga penelitian ini bermanfaat pada pimpinan-pimpinan pada perusahaan agar dapat memastikan tingkat keberhasilan Pegawai dalam menjalankan tugas serta tanggung jawab dalam suatu pekerjaan. Sampel yang digunakan sebanyak 35 responden yang ditetapkan dengan menggunakan metode Michine Learning. Teknik

penarikan sampel yang digunakan adalah dengan metode Algoritma C4.5. Kriteria responden yang dipilih adalah Pegawai tetap pada level main office. Pengolahan dan Analisis Data diolah dengan menggunakan Michine Learning.

### Pengumpulan Data

Adapun tingkat pemberian frekuensi atau angka dalam menentukan frekuensi tingkat kepuasan kerja dan tingkat kedisiplinan yang mencakup penentuan nilai Sedang, Tinggi, Rendah, dan penentuan nilai puas dan tidak puas pada atribut- atribut kepuasan Pegawai yaitu:

**Tabel 1. Frekuensi Pemberian Nilai Atribut**

|            | Tingkat Kepuasan Kerja |        |        |
|------------|------------------------|--------|--------|
| Keterangan | Tinggi                 | Sedang | Rendah |
| Frekuensi  | 75-100                 | 50-75  | 0-50   |

|                 | Lama Bekerja |
|-----------------|--------------|
| Pegawai Tetap   | 10 – 20 dst  |
| Pegawai Kontrak | 6 – 10 Thn   |

### Persiapan Data

**Tabel 2. Atribut Penilaian Kepuasan Pegawai**

| Nama Atribut       | Deskripsi   |
|--------------------|---|
| Nama Responden     | Nama individu yang menjadi responden dalam pengumpulan data Pegawai |
| Lama Bekerja       | Lama Bekerja responden pada perusahaan                              |
| Tingkat Pendidikan | Tahapan yang dimiliki responden                                     |
| Upah               | Ukuran dari hasil kerja yang di capai responden                     |
| Kenyamanan         | Lingkungan Kerja Yang Baik, Sangat Baik dan Kurang Baik             |

### Pemilihan Model *Mechine Learning*

Metode algoritma C4.5 telah di pilih sebagai metode yang sesuai penelitian ini, algoritma C4.5 merupakan algoritma dalam machine learning yang mengandalkan konsep klasifikasi yang merupakan pengembangan dari Algoritma ID3.

Pemilihan algoritma C4.5 dalam kasus ini memiliki alasan yang kuat, tingkat optimisasi penilaian Kepuasan Pegawai cenderung bergantung factor-factor terkait mengenai Kepuasan Pegawai, sebagai seorang peneliti penting untuk memahami kekuatan serta kelemahan masing-masing algoritma machine learning serta kemampuannya dalam menangani data yang diberikan, dengan pemahaman yang mendalam mengenai metode algoritma C4.5 dengan masalah yang dihadapi, pemilihan model ini diharapkan dapat menghasilkan prediksi yang baik mengenai optimisasi penilaian Kepuasan Pegawai pada Dinas Pangan Estate yang bermanfaat bagi penelitian mengenai penilaian.

### Pelatihan Model

Analisa pelatihan metode Pada tahapan analisa metode dilakukan analisa terhadap metode yang digunakan dalam penelitian ini yaitu algoritma C4.5. analisa algoritma C4.5 yaitu membangun pohon keputusan berdasarkan nilai gain tertinggi. tahapan penerapan algoritma C4.5 adalah sebagai berikut:

1. Menghitung jumlah kasus dari masing-masing atribut
2. Menghitung nilai entropy dari masing-masing kasus dengan menggunakan persamaan (2.1).
3. Menghitung nilai gain dari masing-masing atribut dengan menggunakan persamaan (2.2).
4. Memilih atribut sebagai akar berdasarkan dari atribut yang memiliki nilai gain yang tertinggi.
5. Mengulangi perhitungan entropy dan gain hingga atribut tidak

memiliki cabang kasus sehingga terbentuknya pohon keputusan.

Rumus Menghitung Entropy

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

Rumus Menghitung Nilai Gain

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

Berdasarkan rumus dasar entropy dan rumus dasar gain di atas dapat di jabarkan dengan menggunakan rumus entropy dan gain dengan memasukkan atribut- atribut yang sudah ditetapkan sebagai optimisasi penilaian Kepuasan Pegawai sebagai berikut:

Rumus Entropy

$$= ((-Jumlah\ Ya/All\ Total) * \ln \log_2(Jumlah\ Ya/All\ Total) + (-Jumlah\ Tidak/All\ Total) * \ln \log_2(Jumlah\ Tidak/All\ Total))$$

$$= (Entropy\ Total) - ((Jml\ tahun/All\ Total) * Entropy\ Tahun\ Ke\ 1) - ((Jml\ Thn\ Ke\ 2/All\ Total) * Entropy\ Tahun\ Ke\ 2)$$

klasifikasi dalam pemberian status penilaian dengan mengembangkan sebuah model analisis menggunakan Machine Learning (ML). Algoritma C4.5 mampu melakukan clusterisasi terhadap pengelompokan data Pegawai. Kepuasan Algoritma C4.5 juga dapat berkontribusi untuk memberikan hasil prediksi yang tepat dan akurat, adapun variable analisa yakni, nama Pegawai, lama bekerja, tingkat pendidikan, tingkat kepuasan kerja, tingkat kedisiplinan. Hasil penelitian ini menyajikan bahwa algoritma C4.5 memberikan ketepatan dengan akurat dalam bentuk pohon keputusan status optimisasi penilaian, dengan hasil tersebut maka penelitian secara keseluruhan dapat memberikan kontribusi bagi pihak perusahaan dalam optimisasi terkhusus dalam analisis penilaian Kepuasan Pegawai, penelitian ini perlu di lakukan untuk member kemudahan dalam member penilaian yang tepat sesuai dengan ketentuan prestasi Pegawai.

**Tabel 3. Perumusan Dengan Nilai Entropy Dan Gain**

| Mencari nilai entropy dan gain |                 |       |                |           |              |             |             |
|--------------------------------|-----------------|-------|----------------|-----------|--------------|-------------|-------------|
| NODE                           | keterangan      | total | main responden | main (ya) | main (tidak) | entropy     | gain        |
| 1                              | total           | 35    | 23             | 12        |              | 0,927526588 |             |
| 2                              | lama bekerja    |       |                |           |              |             | 0,670454567 |
|                                | Pegawai tetap   | 22    | 22             | 0         | 0            |             |             |
|                                | Pegawai kontrak | 13    | 3              | 11        |              | 0,692116981 |             |
| 3                              | Pendidikan      |       |                |           |              |             | -0,01948564 |
|                                | S1              | 18    | 12             | 6         |              | 0,918295834 |             |
|                                | D3              | 17    | 10             | 7         |              | 0,977417818 |             |
| 4                              | Upah            |       |                |           |              |             | 0,270229354 |
|                                | Tinggi          | 17    | 14             | 3         |              | 0,672294817 |             |
|                                | Sedang          | 13    | 9              | 4         |              | 0,89049164  |             |
|                                | Rendah          | 5     | 0              | 5         |              | 0           |             |
| 5                              | Kenyamanan      |       |                |           |              |             | 0,435671792 |
|                                | Sangat Baik     | 14    | 10             | 4         |              | 0,863120569 |             |
|                                | Baik            | 17    | 12             | 5         |              | 0,184189995 |             |
|                                | Kurang Baik     | 2     | 1              | 1         |              | 1           |             |
|                                | Cukup           | 2     | 0              | 2         |              | 0           |             |

### Analisa Model Machine Learning

Analisa bertujuan untuk melakukan proses prediksi dan

### Perhitungan Manual

Menggunakan algoritma C4.5 dengan jumlah Pegawai sebanyak 35 sampel responden, sampel data analisis Pegawai dengan atribut sebagai berikut:

- Nama Pegawai : 35 responden
- Lama Bekerja : 1-15 Thn
- Tingkat Pendidikan : D3-S1
- Upah : Tinggi, Sedang, Rendah
- Tingkat Kenyamanan : Sangat baik, Baik, Cukupbaik, Kurang baik

Langkah – langkah perhitungan manual algoritma C4.5 adalah sebagai berikut:

1. Menghitung jumlah kasus dari masing –masing atribut
2. Menghitung nilai entropy dari masing-masing dengan menggunakan persamaan (2.1)
3. Menghitung nilai gain dari masing – masing atribut dengan menggunakan persamaan (2.2 )
4. Memilih atribut sebagai akar berdasarkan dari atribut yang memiliki nilai lain yang tertinggi.
5. Mengulangi perhitungan entropy dan gain hingga atribut tidak memiliki cabang kasus sehingga terbentuk nya pohon keputusan.

### Penggunaan Alat Bantu Menggunakan Aplikasi

Dalam hal ini penggunaan alat bantu atau aplikasi untuk menentukan hasil dari perhitungan entropy dan gain yaitu penggunaan aplikasi Excel sebagai berikut:

**Tabel 4. Pengujian Penentuan Nilai Entropy Dan Gain**

| Mencari Nilai Entropy dan Gain |                 |       |                  |           |             |             |
|--------------------------------|-----------------|-------|------------------|-----------|-------------|-------------|
| NODE                           | keterangan      | total | main             | main      | entropy     | gain        |
|                                |                 |       | responden ( ya ) | ( tidak ) |             |             |
| 1                              | Total           | 35    | 23               | 12        | 0,927526388 |             |
| 2                              | lama bekerja    |       |                  |           |             | 0,670454567 |
|                                | Pegawai tetap   | 22    | 22               | 0         | 0           |             |
|                                | Pegawai kontrak | 13    | 3                | 11        | 0,692116981 |             |
| 3                              | Pendidikan      |       |                  |           |             | -0,01948564 |
|                                | S1              | 18    | 12               | 6         | 0,918295834 |             |
|                                | D3              | 17    | 10               | 7         | 0,977417818 |             |
| 4                              | Upah            |       |                  |           |             | 0,270229354 |
|                                | Tinggi          | 17    | 14               | 3         | 0,672294817 |             |
|                                | Sedang          | 13    | 9                | 4         | 0,89049164  |             |
|                                | Rendah          | 5     | 0                | 5         | 0           |             |
| 5                              | Kenyamanan      |       |                  |           |             | 0,435671792 |
|                                | Sangat Baik     | 14    | 10               | 4         | 0,863120569 |             |
|                                | Baik            | 17    | 12               | 5         | 0,184189995 |             |
|                                | Kurang Baik     | 2     | 1                | 1         | 1           |             |
|                                | Cukup           | 2     | 0                | 2         | 0           |             |

### Pengambilan Nilai Gain Tertinggi

Tabel di atas menunjukkan perhitungan node 2 dengan nilai gain tertinggi ada pada data atribut lama bekerja yaitu: 0,670454567 maka

atribut lama bekerja akan menjadi akar selanjutnya, kemudian perhitungan selanjutnya akan di lanjutkan pada perhitungan node 3.

**Tabel 5. Nilai Gain Tertinggi**

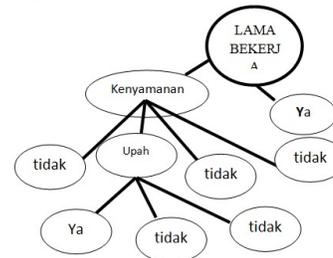
|   |              |                 |    |    |    |             |             |
|---|--------------|-----------------|----|----|----|-------------|-------------|
| 1 | Lama Bekerja |                 |    |    |    |             | 0,670454567 |
|   |              | Pegawai tetap   | 22 | 22 | 0  | 0           |             |
|   |              | Pegawai kontrak | 13 | 3  | 11 | 0,692116981 |             |

**Tabel 6. Pegawai Tetap**

| Pegawai Tetap |            |        |             |          |
|---------------|------------|--------|-------------|----------|
| node          | pendidikan | Upah   | Kenyamanan  | Kepuasan |
| 1.            | D3         | Sedang | baik        | tidak    |
| 2.            | D3         | sedang | cukup       | tidak    |
| 3.            | S1         | tinggi | baik        | ya       |
| 4.            | D3         | rendah | baik        | tidak    |
| 5.            | S1         | tinggi | sangat baik | tidak    |
| 6.            | S1         | tinggi | baik        | ya       |
| 7.            | D3         | rendah | sangat baik | tidak    |
| 8.            | D3         | tinggi | kurang baik | tidak    |
| 9.            | D3         | Sedang | baik        | tidak    |
| 10.           | S1         | rendah | sangat baik | tidak    |
| 11.           | S1         | rendah | baik        | tidak    |
| 12.           | S1         | Sedang | cukup       | tidak    |
| 13.           | D3         | rendah | kurang baik | tidak    |

### Pohon Keputusan

Berikut merupakan pohon keputusan yang dihasilkan dari perhitungan manual.



**Gambar 2. Pohon Keputusan**

### Ruler

Dari Hasil Pohon Keputusan diatas terdapat ruler :

Jika lama bekerja yang didapat Pegawai tetap maka kepuasan “ ya”

Jika lama bekerja sama Pegawai kontrak dan kenyamanan baik dan upah tinggi maka kepuasan “ya”

Jika lama bekerja sama Pegawai kontrak dan kenyamanan sangat baik maka kepuasan “tidak”

Jika lama bekerja sama Pegawai kontrak dan kenyamanan kurang baik maka kepuasan “tidak”

Jika lama bekerja sama Pegawai kontrak dan kenyamanan cukup maka kepuasan “tidak”

Jika lama bekerja sama Pegawai kontrak dan kenyamanna baik dan upah sedang maka kepuasan “tidak”

Jika lama bekerja sama Pegawai kontrak dan kenyamanan baik dan upah rendah maka kepuasan “tidak”

### Peroses Evaluasi Penilaian Kepuasan Pegawai.

Evaluasi dan Validasi Setelah data diuji menggunakan Matlab, maka akan terbentuk pohon keputusan dari data tersebut. Kemudian akan dilakukan penghitungan akurasi, precision, dan recall, selanjutnya dilakukan evaluasi dan validasi hasil dihitung menggunakan rumus akurasi, precision dan recall sebagai berikut.

- Menguji metode/ model menggunakan data testing. Rumus untuk menghitung akurasi adalah sebagai berikut:

$$Akurasi = \frac{\text{Jumlah yang diklasifikasikan benar}}{\text{total sampel testing yang di uji}} \quad (1)$$

- Menghitung precision dan recall untuk mengetahui akurasinya. Rumus Precision dan Recall adalah sebagai berikut:

$$P = \frac{TP}{TP+FP}$$

$$r = \frac{TP}{TP+FN} \quad (2)$$

Keterangan:

p = Precision r = Recall

TP = True Positive

FN = False Negative

### Pengujian Rules Terhadap Data Pegawai

Pengujian terhadap validitas sistem bertujuan untuk mengetahui apakah solusi yang dihasilkan oleh pohon keputusan tersebut valid atau tidak. Rules tersebut dikatakan valid jika jumlah dan nama Pegawai yang puas sesuai dengan data set, yang mana keseluruhan data set keseluruhan berjumlah 500-1000 responden yang mana data training yang diambil sebanyak 35 data Pegawai yang digunakan dalam pengujian.

Dimana dari 35 data di atas mendapatkan hasil evaluasi untuk dilakukan pengujian sebagai berikut:

| Prediksi | Ground truth |       |
|----------|--------------|-------|
|          | Ya           | Tidak |
| Ya       | 23           | 12    |
| tidak    | 12           | 23    |

Tabel data diatas merupakan data prediksi yang akan implementasikan menggunakan akurasi dan recall yang merupakan dari 35 sampel data yang terdapat error sebanyak 12 data, data bernilai positif atau benar sebanyak 32 .

### Evaluasi Model Pengujian Pertama

Perhitungan akurasi merupakan pengujian pertama yang dilakukan dengan cara membagi jumlah data yang diklasifikasikan secara benar dengan total sampel data testing yang di uji.

$$Akurasi = \frac{\text{Data Testing} - \text{Eror}}{\text{Jumlah Data Testing}}$$

$$= \frac{35 - 12}{35}$$

$$= 0,65714285$$

$$= 65 \%$$

### Evaluasi Model Pengujian Kedua

Precision merupakan jumlah sampel berkategori positif diklasifikasikan benar dibagi dengan

total sampel yang akan diklasifikasikan sebagai sampel positif. Menghitung nilai precision dihitung dengan cara membagi jumlah data benar yang bernilai positif ( True Positive ) dibagi dengan jumlah data benar yang bernilai positif ( True Positive ) dan data salah yang bernilai positif ( False Negative)

$$\text{Precision} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{32+32}{32+32+12+12} = 0,727272 = 72\%$$

### Evaluasi model pengujian ketiga

Recall merupakan jumlah sampel diklasifikasikan positif dibagi dengan total sampel dalam testing set berkategori positif. Recall dihitung dengan cara membagi data benar yang bernilai positive ( True Positive ) dengan hasil penjumlahan dari data yang benar yang bernilai positif ( True Positive ) dan data salah yang bernilai negative ( False Negative).

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{32}{32+12} = 0,727272 = 72\%$$

### Pengolahan Data dan Implementasi

Pada tahapan ini dilakukan pengolahan data dan tahapan Implementasi menggunakan aplikasi Orange 3.35.2 dengan menggunakan data yang sudah diolah sebelumnya. Pada implementasi menggunakan data testing dan data training. Data training berisi 80% dari data yang ada dan data testing 20% dari data yang ada. Berikut merupakan data training yang akan dimasukkan ke dalam aplikasi. Data training dapat dilihat pada tabel 7.

**Tabel 7. Data Training C4.5**

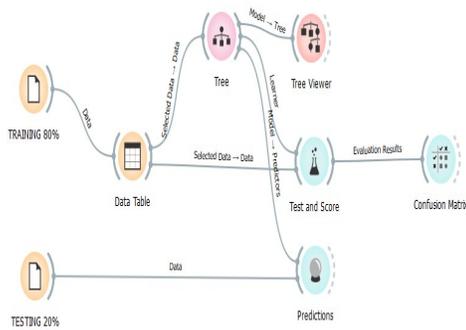
| No  | lama bekerja    | Pendidikan | Upah   | Kenyamanan  | Kepuasan |
|-----|-----------------|------------|--------|-------------|----------|
| 1.  | Pegawai tetap   | D3         | tinggi | sangat baik | ya       |
| 2.  | Pegawai tetap   | S1         | tinggi | baik        | ya       |
| 3.  | Pegawai tetap   | D3         | tinggi | sangat baik | ya       |
| 4.  | Pegawai tetap   | D3         | tinggi | sangat baik | ya       |
| 5.  | Pegawai tetap   | D3         | Sedang | baik        | ya       |
| 6.  | Pegawai tetap   | S1         | tinggi | sangat baik | ya       |
| 7.  | Pegawai tetap   | D3         | tinggi | sangat baik | ya       |
| 8.  | Pegawai kontrak | D3         | rendah | baik        | tidak    |
| 9.  | Pegawai tetap   | S1         | Sedang | baik        | ya       |
| 10. | Pegawai kontrak | S1         | tinggi | sangat baik | tidak    |
| 11. | Pegawai kontrak | S1         | tinggi | baik        | ya       |
| 12. | Pegawai tetap   | D3         | Sedang | sangat baik | ya       |
| 13. | Pegawai kontrak | D3         | rendah | sangat baik | tidak    |
| 14. | Pegawai tetap   | S1         | tinggi | sangat baik | ya       |
| 15. | Pegawai tetap   | S1         | tinggi | baik        | ya       |
| 16. | Pegawai kontrak | D3         | tinggi | kurang baik | ya       |
| 17. | Pegawai tetap   | S1         | Sedang | baik        | ya       |
| 18. | Pegawai kontrak | D3         | Sedang | baik        | tidak    |
| 19. | Pegawai tetap   | D3         | Sedang | baik        | ya       |
| 20. | Pegawai kontrak | S1         | rendah | sangat baik | tidak    |
| 21. | Pegawai kontrak | S1         | rendah | baik        | tidak    |
| 22. | Pegawai tetap   | S1         | Sedang | baik        | ya       |
| 23. | Pegawai tetap   | D3         | tinggi | baik        | ya       |
| 24. | Pegawai tetap   | S1         | tinggi | baik        | ya       |
| 25. | Pegawai tetap   | S1         | tinggi | sangat baik | ya       |
| 26. | Pegawai kontrak | S1         | Sedang | cukup       | tidak    |
| 27. | Pegawai tetap   | S1         | Sedang | baik        | ya       |
| 28. | Pegawai kontrak | D3         | rendah | kurang baik | tidak    |

Data testing dapat dilihat pada tabel 8

**Tabel 8. Data Testing C4.5**

| No | Lama Bekerja    | Pendidikan | Upah   | Kenyamanan  | Kepuasan |
|----|-----------------|------------|--------|-------------|----------|
| 1. | Pegawai kontrak | D3         | Sedang | baik        | tidak    |
| 2. | Pegawai kontrak | D3         | sedang | cukup       | tidak    |
| 3. | Pegawai tetap   | S1         | tinggi | baik        | ya       |
| 4. | Pegawai kontrak | S1         | tinggi | baik        | tidak    |
| 5. | Pegawai tetap   | D3         | tinggi | sangat baik | ya       |
| 6. | Pegawai tetap   | S1         | Sedang | sangat baik | ya       |
| 7. | Pegawai tetap   | D3         | Sedang | sangat baik | ya       |

Secara keseluruhan proses yang dilakukan dengan menggunakan Orange adalah sebagai berikut :



**Gambar 3. Proses Algoritma C4.5**

Proses data mining pada penelitian ini yang pertama adalah menentukan data training dan data testing pada dataset yang digunakan dimana data training yang digunakan pada penelitian ini adalah sebesar 28 data Pegawai dan data testing yang digunakan adalah sebanyak 7 data Pegawai dari 35 total dataset yang ada. Pada gambar diatas ditunjukkan pada widget “TRAINING 80%” dan “TESTING 20%”.

| Evaluation results for target (None, show average over classes) |       |       |       |       |        |       |
|---|-------|-------|-------|-------|--------|-------|
| Model   | AUC   | CA    | F1    | Prec  | Recall | MCC   |
| Tree  | 0.975 | 0.857 | 0.857 | 0.857 | 0.857  | 0.650 |

**Gambar 4. Hasil Widget Test and Score**

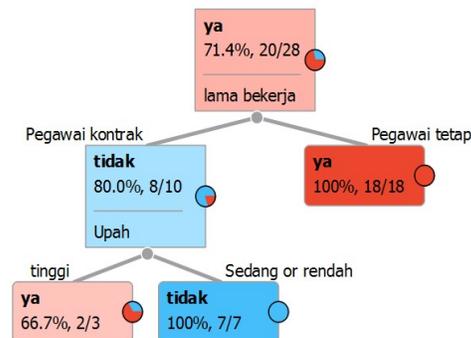
Hasil dari proses pengujian akurasi algoritma C4.5 menggunakan widget “Test and Score” pada aplikasi orange didapatkan akurasi sebesar 85.7 %, F1 Score sebesar 85.7%, precision sebesar 85.7 %, recall 85.7 %, specificity 65%, dan memperoleh nilai AUC (Area Under Curve) sebesar 92.5%. Nilai AUC adalah nilai perhitungan performa suatu indeks data klasifikasi semakin besar nilai AUC

maka semakin baik performa indeks data yang digunakan. Dengan besar nilai AUC sebesar 0.925 atau 92.5% menunjukkan bahwa performa indeks data klasifikasi yang digunakan tersebut termasuk Good Classification

|        |       | Predicted |    |    |
|--------|-------|-----------|----|----|
|        |       | tidak     | ya | Σ  |
| Actual | tidak | 6         | 2  | 8  |
|        | ya    | 2         | 18 | 20 |
| Σ      |       | 8         | 20 | 28 |

**Gambar 5. Confusion Matrix**

Berdasarkan gambar diatas, dari 35 data yang diklasifikasikan dengan algoritma decision tree C4.5. Terdapat 8 data yang dikategorikan Pegawai yang tidak direkomendasikan dan terdapat 20 data Pegawai yang direkomendasikan. 6 data Pegawai sudah memiliki klasifikasi yang tidak sesuai dalam kategori loyalitas Pegawai sedangkan 2 data Pegawai kategori tidak direkomendasikan terprediksi Pegawai yang direkomendasikan. Terdapat 2 data Pegawai yang berkategori loyalitas Pegawai yang terprediksi direkomendasikan sedangkan 18 data Pegawai direkomendasikan yang telah sesuai dengan klasifikasi.



**Gambar 6. Decision Tree Menggunakan Orange**

Berdasarkan gambar diatas, interpretasi dari hasil decision tree dengan algoritma C4.5 adalah sebagai berikut:

1. If Lama Bekerja = Pegawai Kontrak, then Status = Tidak
2. If Lama Bekerja = Pegawai Tetap, then Status = Ya
3. If Pegawai Kontrak, Upah = Tinggi, then status = Ya
4. If Pegawai Kontrak, Upah = Sedang Or Rendah, then Status = Tidak

## V. Kesimpulan Dan Saran

### Kesimpulan

Pada implementasi menggunakan data testing dan data training. Data training berisi 80% dari data yang ada dan data testing 20% dari data yang ada. Dari hasil yang sudah didapatkan maka disimpulkan Hasil dari proses pengujian akurasi algoritma C4.5 menggunakan widget "Test and Score" pada aplikasi orange didapatkan akurasi sebesar 85.7 %, F1 Score sebesar 85.7%, precision sebesar 85.7 %, recall 85.7 %, specificity 65%, dan memperoleh nilai AUC (Area Under Curve) sebesar 92.5%. Dengan besar nilai AUC sebesar 0.925 atau 92.5% menunjukkan bahwa performa indeks data klasifikasi yang digunakan tersebut termasuk Good Classification.

### Saran

Dengan adanya penelitian ini diharapkan kepada para Pegawai Dinas Pangan untuk lebih meningkatkan Kepuasan yang baik serta kerjasama yang baik antar tim. Kepada peneliti, agar lebih banyak menjangkau data untuk sample penelitian sehingga

penelitian ini boleh dipakai oleh banyak perusahaan.

## VI. Daftar Pustaka

- A. Esteban, A. Zafra, and S. Ventura, *Data mining in predictive maintenance systems: A taxonomy and systematic review*, no. March. 2022. doi: 10.1002/widm.1471.
- A. S. A. & R. A. H. & J. k. A. & Z. T. A. & A. A. Z. & B. B. Zaidan *et al.*, "Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review," *J. Med. Syst.*, 2020.
- A. Septiarini, Rizqi Saputra, Andi Tejawati, and Masna Wati, "Deteksi Sarung Samarinda Menggunakan Metode C4.5 Berbasis Pengolahan Citra," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 927–935, 2021, doi: 10.29207/resti.v5i5.3435.
- A. Yudhana, I. Riadi, and M. R. Djou, "Pengembangan Layanan Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1062, 2022, doi: 10.30865/jurikom.v9i4.4515.
- Dr. V. Suma, "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics," *J. Soft Comput. Paradig.*, vol. 2, no. 3, pp. 153–159, 2020, doi: 10.36548/jscp.2020.3.002.
- E. Rasywir, R. Sinaga, and Y. Pratama,

- “Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network (CNN),” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 117–123, 2020, doi: 10.31294/p.v22i2.8907.
- H. M. Siregar, “Implementasi Metode C4.5 Pada Perancangan Aplikasi Sistem Pakar Diagnosa Bronkiektasis,” *Bull. Inf. Technol.*, vol. 1, no. 3, p. 6, 2020.
- I. P. Ninditama, I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, “Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang,” vol. 15, no. 2, pp. 37–49, 2020.
- L. J. Muhammad, M. M. Islam, S. S. Usman, and S. I. Ayon, “Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients’ Recovery,” *SN Comput. Sci.*, vol. 1, no. 4, pp. 1–7, 2020, doi: 10.1007/s42979-020-00216-w.
- N. K. Sriwinarti and P. Juniarti, “Analisis Metode K-Nearest Neighbors ( K-NN ) Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa ( Analysis of K-Nearest Neighbors ( K-NN ) and C4.5 Methods in Predicting Student Graduation ),” vol. 3, no. 2, pp. 106–112, 2021.
- R. Aulia *et al.*, “Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode C4.5,” pp. 106–111, 2020.
- R. Retnosari, “Analisa Kelayakan Kredit Usaha Mikro Berjalan pada Perbankan dengan Metode C4.5,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 8, no. 1, pp. 53–59, 2021, [Online]. Available: <https://ejournal.lppmunsera.org/index.php/PROSISKO/article/view/2848>
- S. AriePrayoga, I. Nawangsih, and T. NgudiWiyatno, “Implementasi Metode Naïve Bayes Classifier Untuk Identifikasi Jenis Jamur,” *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 2, pp. 134–144, 2019.
- Y. Cui, “Intelligent Recommendation System Based on Mathematical Modeling in Personalized Data Mining,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6672036.