Prediksi Pengunduran Diri Calon Karyawan Tetap PT. Asia Pasific Fiber Menggunakan Algoritma Decision Tree Berbasis Particle Swarm Optimization Selama Masa Pandemi

**Tasya Agisti Tioria1, Yusran Timur Samuel2**

1,2Fakultas Teknologi Informasi, Universitas Advent Indonesia

e-mail: 11981007@unai.edu, 2y.tarihoran@unai.edu

**Abstrak**

PT. Asia Pasific Fiber merupakan perusahaan berlokasi di Karawang yang bergerak dalam bidang tekstil. Seperti perusahaan pada umumnya, tidak sedikit karyawan tetap yang memutuskan untuk berhenti walaupun perusahaan sudah memberikan tambahan fasilitas, gaji dan pesangon. Pengunduran diri karyawan mengharuskan perusahaan tetap berjalan walaupun kuantitas dan efisiensi perusahaan sedang terganggu. Maka untuk meningkatkan ketepatan penentuan karyawan tetap, penulis menerapkan teknik *data mining* C4.5 berbasis PSO untuk mengetahui pengaruh atribut dalam pengunduran diri karyawan. Pemilihan ini menggunakan data karyawan PT. Asia Pasific Fiber selama bulan Januari 2020 hingga Agustus 2022 dengan atribut usia, jenis kelamin, status menikah, jabatan, gaji, lama kerja, pendidikan, evaluasi kerja, persentase kehadiran karyawan dan mengundurkan diri sebagai atribut kelas. Peneliti menggunakan Rapidminer dalam memprediksi pengunduran diri calon karyawan PT. Asia Pasific Fiber. Data yang diperoleh dari perusahaan dilakukan *preprocessing* yaitu *data cleaning* dan *data transformation.* Penulis menggunakan *Attribute Weight,* *accuracy*, *precision*, dan *recall dari Confusion Matrix* sebagai evaluasi hasil akhir. Tabel *Attribute Weight* dari hasil PSO menunjukkan bahwa usia, gaji, evaluasi kerja, dan status menikah merupakan atribut dengan pengaruh tertinggi dalam pengunduran diri karyawan. Dengan hasil prediksi pengunduran diri karyawan adalah *accuracy* 96.50%, *precision* 98.51% dan *recall* 97.78%

**Kata Kunci:** Data Mining, Decision Tree, C4.5, prediksi, PSO, pengunduran diri.

***Prediction of Resignation of Prospective Permanent Employees of PT. Asia Pacific Fiber Using Particle Swarm Optimization-Based Decision Tree Algorithm During the Pandemic***

***Abstract***

*PT. Asia Pacific Fiber is a company located in Karawang which is engaged in the textile sector. Like companies in general, there are not a few permanent employees who decide to quit even though the company has provided additional facilities, salary and severance pay. Resignation of employees requires the company to continue running even though the quantity and efficiency of the company are being disrupted. So to increase the accuracy of determining permanent employees, the authors apply the PSO-based C4.5 data mining technique to determine the effect of attributes on employee resignation. This selection uses employee data of PT. Asia Pacific Fiber during January 2020 to August 2022 with the attributes of age, gender, marital status, position, salary, length of service, education, job evaluation, percentage of employee attendance and resignation as class attributes. Researchers use Rapidminer in predicting the resignation of prospective employees of PT. Asia Pacific Fiber. The data obtained from the company is preprocessed, namely data cleaning and data transformation. The author uses the Attribute Weight, accuracy, precision, and recall from the Confusion Matrix as an evaluation of the final results. The Attribute Weight table from the PSO results shows that age, salary, job evaluation, and marital status are the attributes with the highest influence on employee resignation. The results of predicting employee resignation are 96.50% accuracy, 98.51% precision and 97.78% recall*

***Keywords:*** *Data Mining, Decision Tree, C4.5, prediction, PSO, resignation.*

# Pendahuluan

PT. Asia Pasific Fiber merupakan perusahaan berlokasi di Karawang yang bergerak dalam bidang tekstil. Perusahaan ini berdiri pada tahun 1984 dengan nama Polysindo Eka Perkasa sebelum akhirnya berubah menjadi Asia Pasific Fiber pada tahun 2009. PT. Asia Pasific Fiber termasuk salah satu produsen poliester terkemuka di dalam negeri dan internasional, termasuk Eropa, Amerika Serikat, Asia, Australia dan Timur Tengah yang saat ini memiliki lebih kurang 1071 karyawan per bulan Agustus 2022. PT. APF merupakan perusahaan pabrik yang membutuhkan tenaga kerja yang terlatih dalam mempersiapkan produk yang terbaik. Maka dari itu, karyawan tetap adalah aset utama perusahaan yang menjadi perencana dan pelaku aktif dari aktivitas organisasi [1].

Seperti perusahaan pada umumnya, tidak sedikit karyawan tetap yang memutuskan untuk berhenti dari perusahaan walaupun perusahaan sudah memberikan tambahan fasilitas, gaji dan pesangon [2] [3]. Karyawan tetap adalah aset utama perusahaan yang menjadi perencana dan pelaku aktif dari aktivitas organisasi. Maka dari itu seleksi karyawan kontrak menjadi karyawan tetap termasuk proses yang penting, guna menghasilkan karyawan yang sesuai dengan kebutuhan perusahaan dalam segi mutu. Dengan adanya seleksi tersebut, tugas dan jalannya aktivitas perusahaan dapat terjamin dengan diperolehnya karyawan yang tepat [1].

Selama masa pandemic bulan Januari 2020 hingga Agustus 2022, terdapat total 529 karyawan yang berhenti bekerja pada PT. Asia Pasific Fiber dan 82 karyawan diantaranya adalah karyawan tetap. 236 karyawan pada tahun 2020, 192 karyawan pada tahun 2021, dan 99 karyawan pada tahun 2022 dengan rata-rata lama kerja karyawan merupakan 5,5 tahun. Pengunduran diri karyawan mengharuskan perusahaan tetap berjalan walaupun kuantitas dan efisiensi perusahaan sedang terganggu. Arus keluar-masuk karyawan mengakibatkan kemunduran bisnis serta peningkatan pengeluaran untuk pemasaran, seleksi, dan pelatihan karyawan baru. Hal tersebut juga mengurangi efisiensi kerja, dan produktivitas perusahaan[4]. Meskipun terdapat karyawan baru yang menggantikan, masalah ini mengakibatkan perusahaan kehilangan karyawan yang berpotensi dan berpengalaman sehingga harus melatih karyawan baru, sehingga membutuhkan waktu serta biaya tambahan [5].

Maka untuk meningkatkan efisiensi dan ketepatan penentuan karyawan tetap, penulis menerapkan teknik *data mining* [1]. Penentuan karyawan ini memerlukan banyak pertimbangan dari berbagai indikator seperti usia, jenis kelamin, status menikah, jabatan, gaji, lama kerja, pendidikan, evaluasi kerja, persentase kehadiran karyawan dan mengundurkan diri sebagai atribut kelas dengan menggunakan data karyawan tetap PT. Asia Pasific Fiber.

Salah satu solusi untuk permasalahan di atas menggunakan metode algoritma C4.5, yaitu melakukan klasifikasi dengan membentuk pohon keputusan *(Decision Tree*). Pada penelitian yang telah dilakukan, seperti penelitian mengenai menentukan kemungkinan pengunduran diri mahasiswa dengan menggunakan Algoritma C4.5 memperoleh akurasi 96.81%, sedangkan pada penelitian mengenai diagnosa penyakit kanker payudara menggunakan optimasi PSO akurasi meningkat dari 95.61% menjadi 96.49% [6], [7].

Berdasarkan penelitian terdahulu yang dilakukan oleh Endang Sri Palupi dan Said Mirza Pahlevi dengan judul Klasifikasi *Opportunity* Menggunakan AlgoritmaC4.5, C4.5 dan *Naïve Bayes* Berbasis *Particle Swarm Optimization*. Hasil penelitian menyatakan bahwa Algoritma C4.5 menghasilkan tingkat akurasi yang lebih tinggi dibanding algoritma lain serta setelah menggunakan *Particle Swarm Optimization* dalam melakukan klasifikasi[8]. Pada penelitian lain dengan judul Kajian Komparasi Penerapan Algoritma C4.5, *Neural Network,* dan *SVM* dengan Teknik *PSO* untuk Pemilihan Karyawan Teladan PT. XYZ memperoleh hasil bahwa Algoritma C4.5 memiliki tingkat akurasi paling tinggi dari tiga metode lainnya [9].

Adapun tujuan dari penelitian ini adalah untuk mengetahui atribut yang paling berpengaruh terhadap pengunduran diri karyawan dan untuk membuat model dalam memprediksi pengunduran diri calon karyawan tetap. Besar harapan penulis penelitian ini dapat membantu perusahan prediksi apakah karyawan tetap akan mengundurkan diri atau tidak. Pada akhir penelitian ini kita akan mendapatkan *Attribute Weight* untuk mengetahui besar pengaruh setiap atribut dan nilai akurasi prediksi berdasarkan tabel *confusion matrix.*

# Metode Penelitian

Pada penelitian ini penulis menggunakan Algoritma C4.5 untuk membentuk pohon keputusan atau dikenal juga sebagai Decision Tree karena sifatnya yang sederhana dan diterima oleh orang awam [10], [2]. Seperti halnya dengan algoritma lain, Algoritma C4.5 memiliki kekurangan dalam menangani data berdimensi sangat tinggi, sehingga membutuhkan waktu yang lebih lama dan kurangnya akurasi prediksi [11]. Maka penulis memutuskan menerapkan algoritma *Particle Swarm Optimization* untuk mengoptimalkan kinerja klasifikasi Algoritma C4.5. Tahapan penelitian dijabarkan melalui bagan pada Gambar 1.

Mulai

Memperoleh Data

Preprocessing Data

Pembuatan Model C4.5 + PSO

Mengevaluasi Model

Menarik Kesimpulan

Selesai

Gambar 1 Diagram penelitian

## Pengumpulan Data

Pengumpulan data dilakukan melalui Riset Lapangan, berupa langsung ke objek untuk memperoleh data primer PT. Asia Pasific Fiber Karawang pada Departemen Sumber Daya Manusia. Data yang diperoleh merupakan data yang berlangsung selama masa pandemi yaitu selama bulan Januari 2020 hingga Agustus 2022. Jumlah populasi penelitian yang diperoleh adalah 1607 data.

## *Attribute*

Berdasarkan penelitian terdahulu, penulis mengambil beberapa atribut dalam prediksi pengunduran diri calon karyawan, sebagai berikut:

Tabel 1 *Attribute*

|  |  |
| --- | --- |
| INDIKATOR | KETERANGAN |
| Usia | Usia karyawan digolongkan menjadi empat yaitu; 20-29, 30-39, 40-49, 50-59. |
| Jenis Kelamin | Pria atau Wanita |
| Status Menikah | Status menikah karyawan serta tanggungannya. TK: tidak menikah, K0: menikah tanpa tanggungan, K1: menikah 1 tanggungan, K2: menikah 2 tanggungan, K3: menikah 3 tanggungan, K4: menikah 4 tanggungan, K5: menikah 5 tanggungan. |
| Jabatan | Jabatan karyawan yaitu GM (General Manager), MGR (Manager), OP (Operator), SS (Senior Superintendent), S (Superintendent), SV (Supervisor). |
| Gaji | Empat golongan gaji karyawan yaitu; <5 juta, 5 juta – 10 juta, 10 juta -20 juta, dan >20 juta. |
| Lama Kerja | Lama karyawan bekerja pada perusahaan hingga berhenti. |
| Pendidikan | Pendidikan karyawan terdiri dari SMP, SMU, D1, D3, S1, S2. |
| Evaluasi Kerja | Grade kualitas kerja karyawan, tiga tahun sebelum pengunduran diri. A bernilai ≥90, B bernilai 80≥89, C bernilai 70≥79, D bernilai <69. |
| Persentase Kehadiran | Persentase kehadiran karyawan masuk kantor. |
| Mengundurkan Diri | Atribut kelas diaman Ya untuk karyawan yang mengundurkan diri dan Tidak untuk karyawan yang tidak mengundurkan diri. |

## Preprocessing Data

Preprocessing data adalah proses menghilangkan noise pada data yang tidak relevan misalnya data yang hilang, data yang tidak valid, data di luar lingkup penelitian. Hasil dari prepocessing menghasilkan 713 data dari data sebelumnya yang sudah diperoleh. Data tersebut dapat ditangani dengan cara dibuang. Proses ini akan mempengaruhi hasil performa dari teknik data mining yang akan digunakan. Data dimasukkan dan diolah menggunakan *Microsoft Office Excel* untuk memperbaiki kesalahan penulisan tahun, menghapus data yang tidak lengkap, menghapus data di luar lingkup tahun penelitian, dan diluar karyawan tetap perusahaan.Penulis melakukan transformasi data umur melalui data tanggal lahir. Penjelasan proses *preprocessing* dapat dilihat dibawah:

* + - 1. *Cleaning*

Pada tahap ini, penulis membersihkan dapat berdasarkan data yang tidak lengkap atau data yang tidak terisi. Jika *attribute* kosong atau tidak bernilai maka data tersebut akan dihapus, dikarenakan penulis tidak memiliki data untuk mengisi *attribute* yang kosong. Terdapat beberapa data yang tidak memiliki nilai persentase kehadiran, maka penulis menghapus data tersebut. Beberapa data yang dihapus dapat dilihat pada tabel 2.

Tabel 2 *Data Cleaning* yang dihapus

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Pria | TK | Operator | <5 juta | 0-5 | SMU | C |  | Tidak |
| 50-59 | Pria | K3 | Superintendent | 10 juta–20 juta | 0-5 | SMU | C |  | Tidak |
| 20-29 | Pria | TK | Operator | <5 juta | 0-5 | SMU | C |  | Tidak |

* + - 1. *Transformation*

Transformasi data adalah pengubahan format data menjadi format yang dibutuhkan. Data Transformation, bentuk awal data merupakan tanggal kelahiran karyawan. Penulis melakukan transformasi dari tanggal lahir menjadi umur karyawan menggunakan rumus *Microsoft Office Excel*, transforasi atribut jabatan menjadi singkatan yaitu GM, MGR, OP, S, SS, SV. Kemudian transformasi gaji menjadi empat golongan. Penjelasan lebih jelas dapat dilihat pada tabel 3, tabel 5 dan tabel 7 data sebelum dilakukan transformasi data, dan hasil transformasi data pada tabel 4, tabel 6, tabel 8.

Tabel 3 dataset sebelum dilakukan *Data Transformation* tanggal lahir

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tgl Lahir | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 9/22/1993 | Wanita | TK | Operator | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 11/20/1968 | Pria | K3 | Superintendent | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 6/18/1987 | Pria | K2 | Operator | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

Tabel 4 dataset setelah dilakukan *Data Transformation* tanggal lahir

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | Operator | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | Superintendent | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 30-39 | Pria | K2 | Operator | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

Tabel 5 dataset sebelum dilakukan *Data Transformation* jabatan

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | Operator | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | Superintendent | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 30-40 | Pria | K2 | Operator | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

Tabel 6 dataset setelah dilakukan *Data Transformation* jabatan

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | OP | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | S | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 30-40 | Pria | K2 | OP | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

Tabel 7 dataset sebelum dilakukan *Data Transformation* gaji

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | Operator | 6,146,550 | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | Superintendent | 12,377,767 | 20-30 | D3 | D | 90-100 | Tidak |
| 30-40 | Pria | K2 | Operator | 4,818,500 | 10-20 | SMU | D | 90-100 | Tidak |

Tabel 8 dataset setelah dilakukan *Data Transformation* gaji

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | Operator | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | Superintendent | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 30-40 | Pria | K2 | Operator | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

## Evaluasi

Pengujian model algoritma Decision Tree C4.5 menggunakan evaluasi *Confusion Matrix*. Evaluasi dilakukan dengan membandingkan hasil tingkat *accuracy*, *precision*, dan *recall* dari hasil prediksi algoritma Decision Tree C4.5 pada *RapidMiner*. *Confusion Matrix* merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. Berdasarkan tabel *Confusion Matrix* pada Tabel 9, perhitungan *accuracy* menggunakan tabel adalah sebagai berikut:

Tabel 9 *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
|  | | *PREDICTED CLASS* | |
| *POSITIVE* | *NEGATIVES* |
| *ACTUAL CLASS* | *POSITIVES* | *True Positive (TP)* | *False Positive (FP)* |
| *NEGATIVES* | *False Negative (FN)* | *True Negative (TN)* |

(1)

(2)

(3)

*Precision* didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Sedangkan *Recall* didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. *Precision* and *Recall* dapat diberi nilai dalam bentuk angka dengan menggunakan perhitungan persentase (1-100%) atau dengan menggunakan bilangan antara 0-1 [12]. Proses formulasi keputusan hasil analisis melalui visualisasi mengenai metode yang digunakan dalam memperoleh pengetahuan [17].

## Data Mining

*Data mining* merupakan ilmu mempelajari metode dalam menganalisa pola atau nilai tambah pengetahuan dari data yang sangat besar mencakup *database, data warehouse, repository, dsb*. Pada umumnya data mining digunakan untuk memperoleh informasi bersifat implisit dan belum diketahui yang terdapat dalam basis data yang besar atau bisa juga disebut Knowledge Discovery in Database (KDD) [13].

## Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diciptakan oleh peneliti kecerdasan buatan bernama J. Rose Quinlan pada tahun 1970-an [14]. Algoritma C4.5 digunakan untuk membentuk pohon keputusan (*Decision Tree*) dalam metode klasifikasi dan prediksi. Pohon keputusan berguna untuk menemukan hubungan antara *variabel input* dengan *variabel target*. Pemilihan atribut dalam Algoritma C4.5 menggunakan ukuran berdasarkan perhitungan nilai *entropi* dan *gain* [15].

Rumus *Entropi* dan *Gain* dapat dituliskan pada persamaan 1 dan 2 dibawah:

(4)

Keterangan:

S : Himpunan kasus

n : Jumlah partisi S

pi : proporsi dari Si terhadap S

(5)

Keterangan:

S : Himpunan kasus

A : Fitur

n : Jumlah partisi A

|Si| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S [18]

Secara umum, tahapan membangun pohon keputusan Algoritma C4.5 adalah:

1. Memilih atribut sebagai akar pohon keputusan
2. membentuk cabang setiap nilai
3. Membagi kasus pada cabang
4. mengulangi proses setiap cabang hingga cabang memiliki kelas yang sama [16].

Tabel 10 Hasil perhitungan *Entropy* dan *Gain*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Node |  | Keterangan | Kasus | Mengundurkan Diri | | Entropy | Gain |
| Ya | Tidak |
| 1 |  |  |  |  |  |  |  |
|  | Total |  | 713 | 75 | 638 | 0.485233 |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | usia |  |  |  |  |  | 0.046933 |
|  |  | 20-29 | 52 | 5 | 47 | 0.456684 |  |
|  |  | 30-39 | 158 | 9 | 149 | 0.315265 |  |
|  |  | 40-49 | 268 | 10 | 258 | 0.229836 |  |
|  |  | 50-59 | 235 | 51 | 184 | 0.754691 |  |
|  |  |  |  |  |  |  |  |
|  | jenis kelamin |  |  |  |  |  | 0.002206 |
|  |  | Pria | 669 | 73 | 596 | 0.497250 |  |
|  |  | Wanita | 44 | 2 | 42 | 0.266765 |  |
|  |  |  |  |  |  |  |  |
|  | status menikah |  |  |  |  |  | 0.005183 |
|  |  | TK | 102 | 9 | 93 | 0.430552 |  |
|  |  | K0 | 19 | 1 | 18 | 0.297472 |  |
|  |  | K1 | 151 | 16 | 135 | 0.487609 |  |
|  |  | K2 | 284 | 27 | 257 | 0.453172 |  |
|  |  | K3 | 140 | 18 | 122 | 0.553507 |  |
|  |  | K4 | 13 | 3 | 10 | 0.779350 |  |
|  |  | K5 | 3 | 1 | 2 | 0.918296 |  |
|  |  | K6 | 1 | 0 | 1 | 0.000000 |  |
|  |  |  |  |  |  |  |  |
|  | jabatan |  |  |  |  |  | 0.011195 |
|  |  | Gen. Mgr | 1 | 0 | 1 | 0.000000 |  |
|  |  | Manager | 18 | 2 | 16 | 0.503258 |  |
|  |  | Operator | 424 | 35 | 389 | 0.411091 |  |
|  |  | Sr.Spdt | 18 | 1 | 17 | 0.309543 |  |
|  |  | Superintendent | 52 | 8 | 44 | 0.619382 |  |
|  |  | Supervisor | 200 | 28 | 172 | 0.584239 |  |
|  |  |  |  |  |  |  |  |
|  | gaji |  |  |  |  |  | 0.037990 |
|  |  | <5 juta | 35 | 13 | 22 | 0.951763 |  |
|  |  | 5 juta -10 juta | 624 | 53 | 571 | 0.419337 |  |
|  |  | 10 juta -20 juta | 39 | 4 | 35 | 0.477071 |  |
|  |  | >20 juta | 15 | 1 | 14 | 0.353359 |  |
|  |  |  |  |  |  |  |  |
|  | pengalaman kerja |  |  |  |  |  | 0.030456 |
|  |  | 0-5 | 73 | 9 | 64 | 0.538738 |  |
|  |  | 5-10 | 51 | 8 | 43 | 0.626751 |  |
|  |  | 10-20 | 208 | 5 | 203 | 0.163551 |  |
|  |  | 20-30 | 376 | 51 | 325 | 0.572701 |  |
|  |  | 30-40 | 5 | 1 | 4 | 0.721928 |  |
|  |  |  |  |  |  |  |  |
|  | pendidikan |  |  |  |  |  | 0.029667 |
|  |  | SMP | 19 | 0 | 19 | 0.000000 |  |
|  |  | SMU | 522 | 53 | 469 | 0.473835 |  |
|  |  | D1 | 12 | 0 | 12 | 0.000000 |  |
|  |  | D3 | 48 | 1 | 47 | 0.146094 |  |
|  |  | S1 | 109 | 18 | 91 | 0.646460 |  |
|  |  | S2 | 3 | 0 | 3 | 0.000000 |  |
|  |  |  |  |  |  |  |  |
|  | evaluasi kerja |  |  |  |  |  | 0.010216 |
|  |  | A | 89 | 4 | 85 | 0.264517 |  |
|  |  | B | 290 | 24 | 266 | 0.411826 |  |
|  |  | C | 322 | 45 | 277 | 0.583588 |  |
|  |  | D | 12 | 2 | 10 | 0.650022 |  |
|  |  |  |  |  |  |  |  |
|  | persentase kehadiran |  |  |  |  |  | 0.004872 |
|  |  | 70-80 | 1 | 1 | 0 | 0.000000 |  |
|  |  | 80-90 | 49 | 4 | 45 | 0.407906 |  |
|  |  | 90-100 | 663 | 70 | 593 | 0.486440 |  |

## Particle Swarm Optimization

*Particle Swarm Optimization* pertama kali di kembangkan tahun 1995 oleh Kennedy dan Eberhart, terinspirasi dan perilaku unggas dan ikan dalam mencari sumber makanan. PSO sudah terbukti efektif dan sering digunakan dalam memecahkan masalah optimasi yang bersifat multidimensi dan multi parameter seperti pada metode klasifikasi [17].

## *PSO* merupakan metode penelitian populasi yang menggunakan populasi (*swarm*) individu (*particle*) yang diperbarui dari setiap iterasi yang dilakukan. Setiap partikel diberikan kecepatan awal dan posisi awal. Partikel akan mengikuti partikel lainnya yang sudah menemukan arah solusi [18]. Dalam membuat partikel mencapai solusi optimumnya, setiap partikel bergerak searah dengan posisi terbaik sebelumnya (*pbest*) dan posisi global terbaik (*gbest*) [19]. PSO menghasilkan table *attribute weight* yang berisi nilai berat masing-masing atribut dalam mempengaruhi hasil prediksi dari model data mining.

Pada algoritma ini, pencarian solusi dilakukan dengan membangkitkan populasi secara *random.* Populasi terdiri dari beberapa partikel yang merupakan presentasi dari solusi permasalahan. Partikel tersebut mencari solusi dengan menyesuaikan *gbest* dan *pbest*. Setelah itu, proses pencarian dilakukan dalam jumlah iterasi tertentu hingga mencapai solusi terbaik.

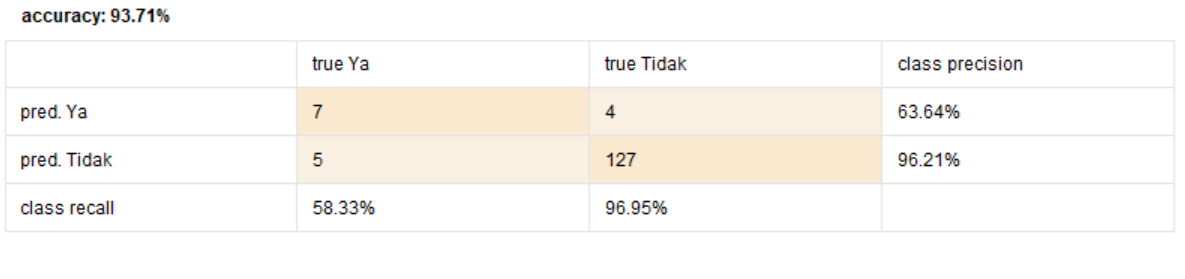
# Hasil

Penelitian ini dilakukan di PT. Asia Pasific Fiber dengan lokasi penelitian di Karawang. Data diperoleh melalui *database* perusahaan dengan batas waktu bulan Januari 2020 hingga Agustus 2022 dengan jumlah data setelah dilakukan preprocessing adalah 713 data. Selanjutnya data dimasukkan dan diolah menggunakan *Microsoft Office Excel* untuk memperbaiki kesalahan penulisan tahun, menghapus data yang tidak lengkap, menghapus data di luar lingkup tahun penelitian. Kemudian melakukan transformasi data umur melalui data tanggal lahir. Adapun hasil *preprocessing* data yang diperoleh adalah sebagai berikut:

Tabel 11 contoh dataset setelah pra-proses

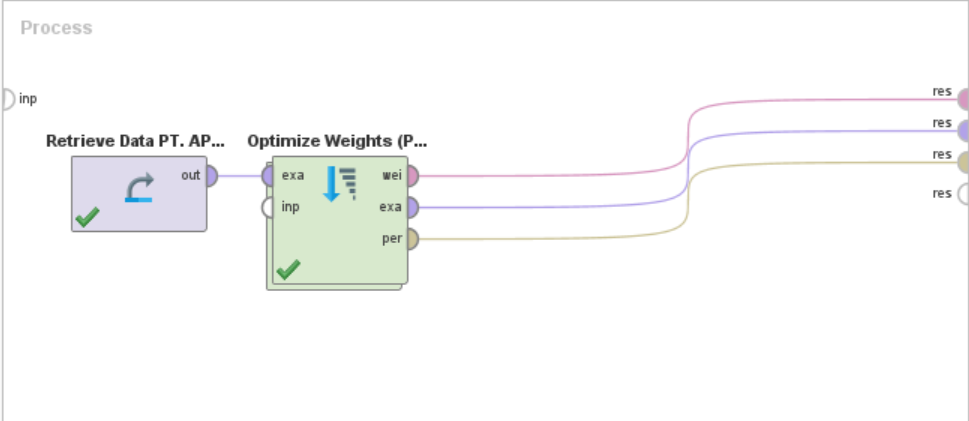
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usia | Jenis Kelamin | Status Menikah | Jabatan | Gaji | Lama kerja | Pendidikan | Evaluasi Kerja | Persentase Kehadiran | Mengundurkan Diri |
| 20-29 | Wanita | TK | Operator | 5 juta-10 juta | 5-10 | SMU | C | 90-100 | Tidak |
| 50-59 | Pria | K3 | Superintendent | 10 juta–20 juta | 20-30 | D3 | D | 90-100 | Tidak |
| 30-40 | Pria | K2 | Operator | <5 juta | 10-20 | SMU | D | 90-100 | Tidak |

Hasil *preprocessing* yang didapat kemudian diolah ke dalam *Decision tree* menggunakan software *Rapidminer* untuk mengetahui prediksi pengunduran diri karyawan. Peneliti melakukan data testing untuk melihat kesesuaian metode yang digunakan dengan kasus yang diteliti, adapun hasil *accuracy* prediksi sebelum menggunakan PSO adalah 93.71%, *precision* 96.21%, *recall* 96.95%.

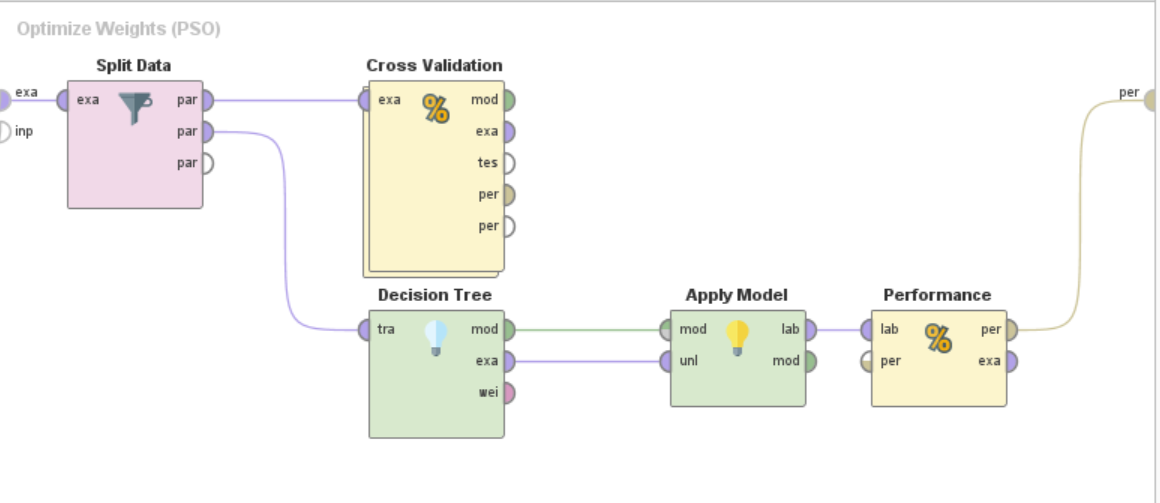


Gambar 2 Hasil akurasi tanpa menggunakan PSO

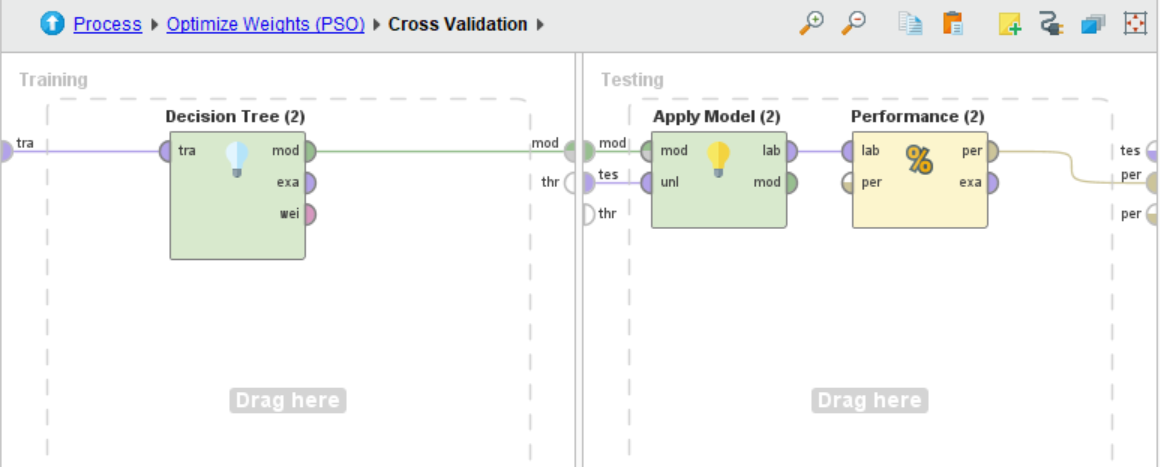
Gambar 3, Gambar 4, Gambar 5merupakan pengujian model Algoritma C4.5 menggunakan aplikasi *Rapidminer.* Penulis menggunakan operator *Retrieve Data* PT. APF pada gambar 3. Penulis menghubungkan dengan operator *Optimize Weights (PSO)* dengan parameter *population size* 5 yaitu jumlah individual/ partikel tiap generasi, *maximum number of generations* 30 yaitu jumlah batas generasi, *inertia weight* 1 yaitu bobot awal, *local best weight* 1 yaitu bobot untuk posisi terbaik partikel, *global best weight* 1 yaitu bobot untuk posisi populasi terbaik, *min weight* 0 yaitu batas rendah bobot dan *max weight* 1 yaitu batas tertinggi bobot. Dalam proses *Optimize Weights (PSO)* penulis menambahkan operator Cross Validation untuk data training dan *Decision Tree C4.5, Apply Model* dan *Performance* untuk data testing dimana sebelumnya diberikan operator *Split Data* untuk memisahkan *data testing* dan *data training* dengan rasio 0.8, 0.2 dan tipe *sampling shuffled sampling.* Pada operator Cross Validation penulis menambahkan operator Decision Tree pada bagian training dan operator Apply Model, Performance pada bagian testing.



Gambar 3 Proses C4.5 menggunakan PSO

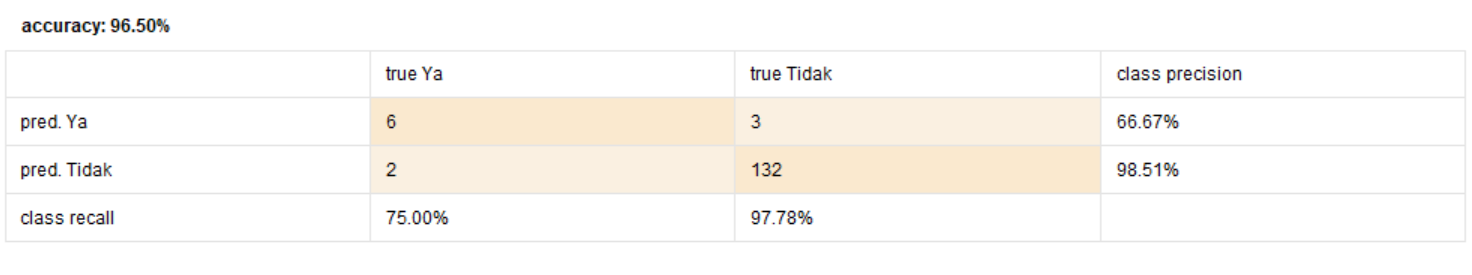


Gambar 4 Proses C4.5 menggunakan PSO

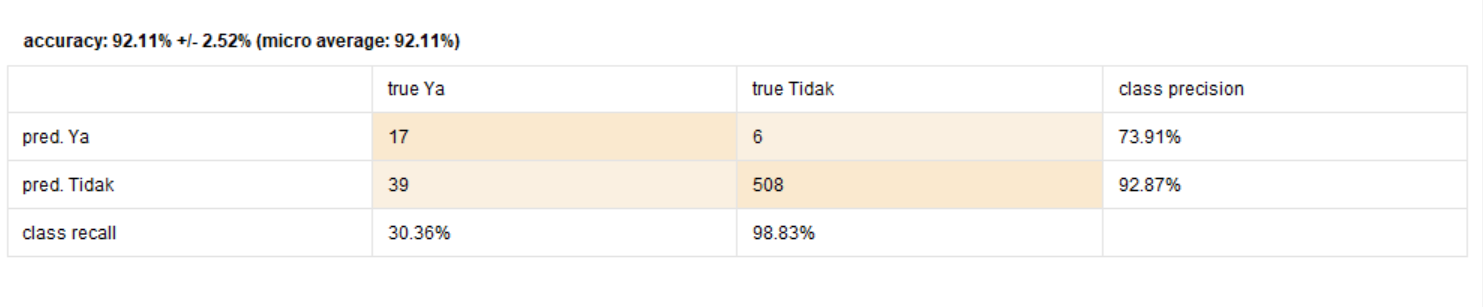


Gambar 5 Proses C4.5 menggunakan PSO

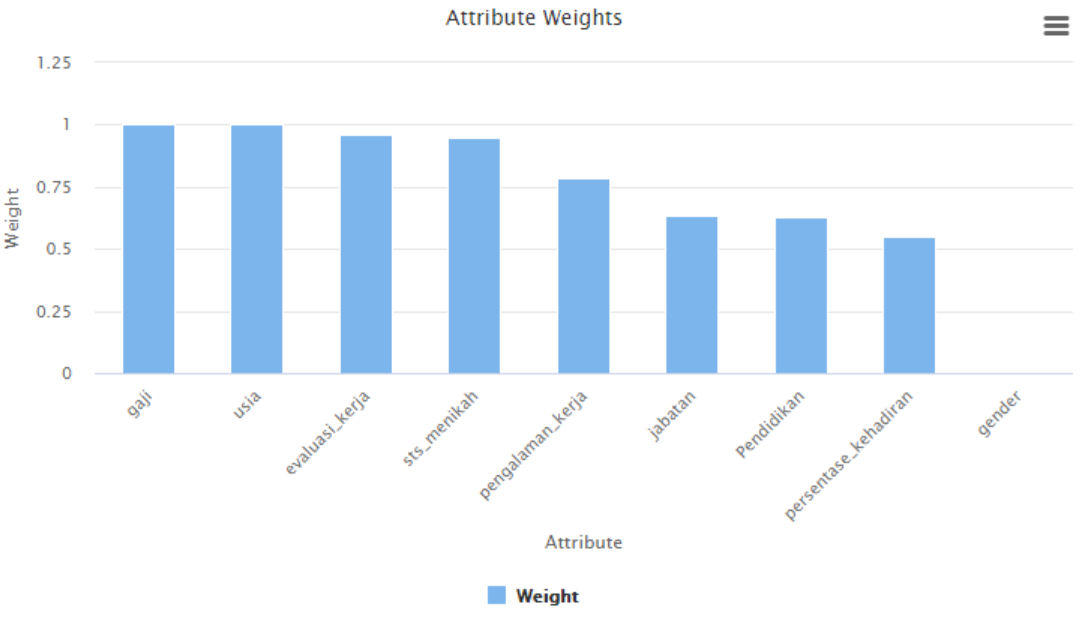
Penulis mendapati hasil Performance data training berupa *accuracy 92.11%* *Precision 92.92%, recall 98.82%*, hasil tersebut cukup baik untuk memprediksi pengunduran calon karyawan tetap, sehingga penulis menerapkan model pada data traning*.* Maka hasil data testing adalah *accuracy* 96.50%, *precision* 98.51% dan *recall* 97.78%. Accuracy adalah ketepatan prediksi pengunduran diri dibandingkan jumlah aktualnya. Precision adalah ketepatan pengunduran diri yang diminta dibandingkan dengan hasil prediksi. Recall adalah prediksi pengunduran diri yang benar dibandingkan dengan jumlah data. Maka model diatas layak untuk digunakan memprediksi pengunduran diri karyawan dan telah mengalami peningkatan sesudah menggunakan PSO, seperti yang tertulis pada Gambar 6.



Gambar 5 Hasil akurasi testing menggunakan PSO

Gambar 6 Hasil akurasi training menggunakan PSO

Berdasarkan tabel *Attribute Weight* didapati bahwa atribut usia, gaji, evaluasi kerja, dan status menikah merupakan empat atribut tertinggi yang kemungkinan mempengaruhi calong karyawan tetap PT. Asia Pasific Fiber untuk mengundurkan diri. Dengan hasil PSO perhitungan berat masing-masing atribut terlihat pada Gambar 7.



**Gambar 7** Attribute Weights hasil PSO

# 

# Gambar 8 Pohon Keputusan Prediksi Pengunduran Diri menggunakan algoritma C4.5

# Dari pohon keputusan tersebut maka diperoleh aturan-aturan atau rule sebagai berikut:

# R1: if Gaji = 10 juta-20 juta then Class = Tidak

# R2: if Gaji = 5 juta-10 juta AND Usia = 20-29 then Class = Tidak

# R3: if Gaji = 5 juta-10 juta AND Usia = 30-39 then Class = Tidak

# R4: if Gaji = 5 juta-10 juta AND Usia = 40-49 then Class = Tidak

# R5: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Operator then Class = Tidak

# R6: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 80-90 then Class = Tidak

# R7: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 90-100 AND Pendidikan = D3 then Class = Tidak

# R8: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 90-100 AND Pendidikan = S1 then Class = Ya

# R9: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 90-100 AND Pendidikan = SMU AND Status Menikah = K1 then Class = Tidak

# R10: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 90-100 AND Pendidikan = SMU AND Status Menikah = K2 then Class = Ya

# R11: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Superintendent AND Persentase Kehadiran = 90-100 AND Pendidikan = SMU AND Status Menikah = K3 then Class = Tidak

# R12: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Supervisor AND Pendidikan = D3 then Class = Tidak

# R13: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Supervisor AND Pendidikan = S1 AND Evaluasi Kerja = A then Class = Tidak

# R14: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Supervisor AND Pendidikan = S1 AND Evaluasi Kerja = B then Class = Tidak

# R15: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Supervisor AND Pendidikan = S1 AND Evaluasi Kerja = C then Class = Ya

# R16: if Gaji = 5 juta-10 juta AND Usia = 50-59 AND Jabatan = Supervisor AND Pendidikan SMU then Class = Tidak

# R17: if Gaji = <5 juta AND Pengalaman Kerja = 0-5 then Class = Tidak

# R18: if Gaji = <5 juta AND Pengalaman Kerja = 10-20 then Class = Tidak

# R19: if Gaji = <5 juta AND Pengalaman Kerja = 20-30 then Class = Ya

# R20: if Gaji = <5 juta AND Pengalaman Kerja = 5-10 then Class = Ya

# R21: if Gaji = >20 juta then Class = Tidak

# Kesimpulan

Dalam penelitian ini dilakukan pengujian model menggunakan metode data mining Algoritma C4.5 dengan menggunakan data karyawan PT. Asia Pasific Fiber di Kota Karawang. Terdiri dari data karyawan selama masa pandemi yaitu selama bulan Januari 2020 hingga bulan Agustus 2022 yaitu sebanyak 713 data. Atribut yang digunakan sebagai parameter prediksi pengunduran diri adalah usia, jenis kelamin, status menikah, jabatan, gaji, lama kerja, pendidikan, evaluasi kerja, persentase kehadiran karyawan dan mengundurkan diri. Data diuji menggunakan RapidMiner yang kemudian menghasilkan nilai *accuracy.* Sebelum menerapkan optimasi PSO, model memperoleh *accuracy* 93.71%, *precision* 96.21%, *recall* 96.95%*.* Setelah menggunakan optimasi PSO, hasil Algoritma C4.5 meningkat menjadi *accuracy* 96.50%, *precision* 98.51% dan *recall* 97.78% dimana lebih baik dari hasil tanpa menggunakan PSO*.* Maka, hasilmetode Algoritma C4.5 berbasis PSO dengan menggunakan aplikasi Rapid Miner dapat menghasilkan informasi untuk mengambil keputusan yang lebih optimal. Dalam hal ini dihasilkan *21 rules* berupa aturan pola dengan atribut usia dan gaji menjadi atribut yang dominan. Dari hasil pohon keputusan menyimpulkan bahwa karyawan yang menerima gaji sebesar 10 juta atau lebih, maka karyawan tidak mengundurkan dari perusahaan. Rules lain juga menunjukkan karyawan dengan gaji 5-10 juta dan berusia 50-59 tahun dengan jabatan Superintendent dan memiliki persentasi kehadiran 80-90 tidak mengundurkan diri. Sedangkan karyawan yang menerima gaji dibawah 5 juta dan memiliki pengalaman bekerja 20-30 tahun atau 5-10 tahun akan mengundurkan diri.

Dari hasil penelitian yang telah dilakukan, masih banyak kekurangan yang bisa diperbaiki. Besar harapan peneliti selanjutnya dapat meningkatkan atribut tambahan yang belum digunakan pada penelitian ini, menerapkan metode data mining lain guna membandingkan akurasi dalam memprediksi pengunduran diri calon karyawan tetap serta memperluas populasi penelitian.

# Daftar Pustaka

[1] Suprapto, Fahrul Fahrezi J.T, dan Edora, “Prediksi Pengangkatan Karyawan Dengan Metode Klasifikasi Algoritma C5. 0 (Studi Kasus CV. T-Pico Jaya Mandiri),” *J. SIGMA*, vol. 13, no. 1, hlm. 35–40, 2022.

[2] Arya Tandy Hermawan, F.X. Ferdinandus, David Boy Tonara, dan Hartarto Junaedi, “Induksi Decision Tree untuk Prediksi Jumlah Buruh yang Berhenti Bekerja,” dipresentasikan pada Konferensi Nasional Inovasi dalam Desain dan Teknologi, Surabaya, 2011. [Daring]. Tersedia pada: https://ideatech.stts.edu/proceeding2011/10-000111\_INF%20Pak%20Arya%20p78-84.pdf

[3] Satriawaty Mallu, “SISTEM PENDUKUNG KEPUTUSAN PENENTUAN KARYAWAN KONTRAK MENJADI KARYAWAN TETAP MENGGUNAKAN METODE TOPSIS,” *J. Ilm. Teknol. Infomasi Terap.*, vol. 1, no. 2, hlm. 36–42, 2015, doi: https://doi.org/10.33197/jitter.vol1.iss2.2015.53.

[4] Max Pangkey, “Analisis Faktor-Faktor Penyebab Pengunduran Diri Karyawan Waktu Tertentu Pada PT. Sinar Pure Foods International,” *J. ILMU Adm. JIA*, vol. 8, no. 3, Des 2012.

[5] Delita Sari dan Susanto, “Mengungkap Tingginya Turnover Intention PT. WBS Semarang,” *Solusi*, vol. 17, no. 2, hlm. 156–177, 2019.

[6] H. Hoiriyah, “Algoritma C4.5 Berbasis Seleksi Atribut untuk Menentukan Kemungkinan Pengunduran Diri Mahasiswa,” *Technol. J. Ilm.*, vol. 9, no. 1, hlm. 67, Jan 2018, doi: 10.31602/tji.v9i1.1104.

[7] M. A. Muslim, S. H. Rukmana, E. Sugiharti, B. Prasetiyo, dan S. Alimah, “Optimization of C4.5 algorithm-based particle swarm optimization for breast cancer diagnosis,” *J. Phys. Conf. Ser.*, vol. 983, hlm. 012063, Mar 2018, doi: 10.1088/1742-6596/983/1/012063.

[8] E. S. Palupi dan S. M. Pahlevi, “KLASIFIKASI OPPORTUNITY MENGGUNAKAN ALGORITMA C4.5, C4.5 DAN NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION,” *INTI Nusa Mandiri*, vol. 14, no. 2, hlm. 233–238, Feb 2020, doi: 10.33480/inti.v14i2.1178.

[9] R. A. Raharjo, “Kajian Komparasi Penerapan Algoritma C4.5, Neural Network, dan SVM dengan Teknik PSO untuk Pemilihan Karyawan Teladan PT. XYZ,” *STRING Satuan Tulisan Ris. Dan Inov. Teknol.*, vol. 1, no. 3, hlm. 345, Apr 2017, doi: 10.30998/string.v1i3.1564.

[10] Bernadus Gunawan Sudarsono dan Alexius Ulan Bani, “C4.5 Prediksi Mahasiswa Berpotensi Berhenti Kuliah Secara Sepihak Menggunakan Data Mining Algoritma C4.5,” *J-SAKTI J. Sains Komput. Dan Inform.*, vol. 4, no. 2, hlm. 359–367, 2020, doi: 10.30645/j-sakti.v4i2.227.

[11] Triowali Rosandy, “Perbandingan Metode Naive Bayes Cassifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan (Study Kasus : KSPPS / BMT AL-FADHILA,” *J. Teknol. Inf. Magister*, vol. 2, no. 01, hlm. 52–62, 2016.

[12] Liliana Swastina, “Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa,” *J. GEMA Aktual.*, vol. 2, no. 1, hlm. 93–98, Mei 2013.

[13] Julkari Sinta Parapat dan Anita Sindar Sinaga, “Data Mining Algoritma C4.5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam,” *J. Ilmu Tek. Elektro Komput. Dan Inform. JITEKI*, vol. 4, no. 2, hlm. 144–154, Des 2018.

[14] Vista Anestiviya dan A. Ferico Octaviansyah Pasaribu, “Analisis Pola Menggunakan Metode C4.5 untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (studi kasus : SMAN 1 NATAR),” *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 1, hlm. 80–85, Mar 2021, doi: https://doi.org/10.33365/jtsi.v2i1.393.

[15] A. Febriani, T. T. Rahmawati, dan E. Sabna, “Implementation of Data Mining to Predict The Feasibility of Blood Donors Using C4.5 Algorithm,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, hlm. 41, Mar 2018, doi: 10.24014/ijaidm.v1i1.4562.

[16] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, dan A. P. Windarto, “Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, hlm. 012024, Jun 2020, doi: 10.1088/1757-899X/874/1/012024.

[17] A. A. Agustian dan A. Bisri, “Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification,” *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 1, Mar 2019, doi: 10.24014/ijaidm.v2i1.6299.

[18] Dinita Rahmalia, “Teknik Penalti pada Optimisasi Berkendala Menggunakan Particle Swarm Optimization,” *JMPM J. Mat. Dan Pendidik. Mat.*, vol. 3, no. 1, hlm. 44–52, Mar 2018, doi: https://doi.org/10.26594/jmpm.v3i1.1071.

[19] Nur Hayatin, Gita Indah Marthasari, dan Lia Nuarini, “Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization,” *J. Online Inform.*, vol. 5, no. 1, hlm. 81–88, 2020, doi: https://doi.org/10.15575/join.v5i1.558.