

PENGEMBANGAN ALGORITMA C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PENENTUAN KELAYAKAN ASURANSI

Jatmiko Indriyanto, Miftakhul Huda dan Ida Afriliana

Universitas Harapan Bangsa, Banyumas Jawa Tengah, Indonesia

Politeknik Harapan Bersama Tegal Jawa Tengah, Indonesia

Politeknik Harapan Bersama Tegal Jawa Tengah, Indonesia

Email: dewajat@gmail.com, mh_iftah@yahoo.com dan idaafriharahap@gmail.com

INFO ARTIKEL

Diterima

25 Februari 2021

Diterima dalam bentuk revisi

15 Maret 2021

Diterima dalam bentuk revisi

Keywords:

C4.5 algorithm; PSO;
insurance

ABSTRACT

The purpose of this study is for the Development of Particle Swarm Optimization-Based C4.5 Algorithm For Determining Insurance Eligibility. logistic regresion methods, decision trees, k-nearest neighbors, naïve bayes and vector machines support. The model will determine or predict the future status of consumers. Similar observations have also been made, but in different ways. In this research, C4.5 classification algorithm based on Particle Swarm Optimization (PSO), the desired accuracy result is better than using only C4.5 algorithm to solve problems in the case of insurance product selection. it can be concluded that the accuracy value obtained in the PSO-based C4.5 algorithm model is 98.93% better when compared to the C4.5 algorithm model of 97.84%. From the results obtained the difference between the two models is worth 0.4%. While for the study using ROC curve for both models is, for the C4.5 algorithm model the AUC value is 0.970 with the Excellent Classification diagnostic sequence, and for the PSO-based C4.5 algorithm model the AUC value is 0.968 with the Excellent Classification diagnostic sequence.

ABSTRAK

Tujuan penelitian ini adalah untuk Pengembangan Algoritma C4.5 Berbasis *Particle Swarm Optimization* Untuk Penentuan Kelayakan Asuransi. metode *logistic regresion, decision trees, k-nearest neighbors, naïve bayes* dan *support vector machines*. Model tersebut akan menentukan atau memprediksi status konsumen dimasa mendatang. Observasi yang mirip juga pernah dilakukan, tetapi dengan cara berbeda. Pada penelitian ini, akan digunakan algoritma klasifikasi C4.5 berbasis *Particle Swarm Optimization* (PSO), hasil ketepatan yang diinginkan lebih bagus dibandingkan hanya memakai algoritma C4.5 untuk mengatasi permasalahan pada

Kata Kunci:

algoritma C4.5; PSO; asuransi

kasus pemilihan produk asuransi. dapat disimpulkan bahwa nilai akurasi yang didapatkan pada model algoritma C4.5 berbasis PSO adalah 98.93% lebih bagus jika dibandingkan dengan model algoritma C4.5 yaitu 97.84%. Dari hasil tersebut didapatkan perbedaan antara kedua model yaitu senilai 0.4%. Selagi untuk penelaahan menggunakan ROC curve bagi kedua model ialah, untuk model algoritma C4.5 nilai AUC adalah 0.970 dengan urutan diagnosa *Excellent Classification*, dan untuk model algoritma C4.5 berbasis PSO nilai AUC adalah 0.968 dengan urutan diagnosa *Excellent Classification*.

Corresponden Author

Email: dewajat@gmail.com

Artikel dengan akses terbuka dibawah lisensi



Pendahuluan

Banyak perusahaan yang bersedia untuk membayar kerugian yang sedikit pada masa sekarang agar terhindar dari risiko kerugian yang besar yang mungkin terjadi di masa datang (Utari & Wibowo, 2020). Karena kemungkinan kerugian di masa datang telah dipindahkan kepada pihak lain yang mau menanggungnya, misalnya perusahaan asuransi (Umam, 2018). Banyak masyarakat tidak menyadari risiko yang akan dihadapi di hari tua, risiko biaya pendidikan bagi anak yang tiap tahunnya selalu bertambah, risiko biaya kesehatan baginya dan keluarga, risiko kendaraan pribadi yang digunakan, risiko bagi keluarga yang ditinggalkan (Maralis & Triyono, 2019).

Asuransi adalah suatu metode untuk meminimalisir risiko dengan jalan memindahkan dan mengkombinasikan ketidakjelasan akan adanya kehilangan keuangan (Indriyanto & Purwanto, 2014). Dengan adanya perusahaan asuransi, kita dapat memindahkan resiko (ketidakpastian) yang kita hadapi menjadi kepastian (Santoso et al., 2018).

Sesuai Undang-Undang Republik Indonesia no.2 tahun 1992 tentang usaha perasuransian, ada pasal 1 ayat 3, “.....dengan tujuan untuk memberikan perlindungan dasar bagi kesejahteraan masyarakat.” yang merupakan tujuan utama dari perusahaan asuransi (Siswanti, 2017). Masyarakat Indonesia pada biasanya, masih tidak mau atau tidak mengindahkan pentingnya perancangan asuransi (Ermawati, 2019). Bila mereka dikunjungi oleh agen asuransi, maka mereka akan mengelak. Agen asuransi seringkali menjual produk yang tidak sesuai dengan kebutuhan kliennya, sehingga klien merasa kehilangan. Agen asuransi juga sering menawarkan produk asuransi, langsung pada produknya, padahal sebagian besar penduduk belum mengerti tentang asuransi (Santoso et al., 2018). Ada agen asuransi yang belum tahu benar mengenai produk yang akan ditawarkan, tetapi langsung terjun ke masyarakat (Rani, 2015). Ada polis asuransi yang

sudah di acc, tetapi ditengah jalan atau pada akhir masa polis bermasalah, dikarenakan informasi yang diberikan tidak tepat. Sehingga menentukan produk asuransi sering tidak akurat (sulistiyanto, 2018).

Metode Penelitian

1. Pengumpulan data

Pada tahap ini diplotkan data yang akan diolah. Mencari data yang tersedia, mendapat data tambahan yang diinginkan, menggabungkan semua data kedalam dataset, termasuk variabel yang diperlukan dalam proses. Hal yang dilakukan dalam pengumpulan data, menentukan poin-poin apa saja data yang akan disimpan dan diperlukan dari nasabah asuransi (Utama, 2013).

2. Pengolahan awal data

Ditahap ini dilakukan pemilihan data, data dibersihkan dan dirubah kebentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model. Hal yang dilakukan dalam pengolahan data, merubah data yang diinput oleh nasabah asuransi, menjadi data yang berguna bagi nasabah asuransi dan perusahaan asuransi (Karlia, 2021).

3. Metode yang diusulkan (*Proposed Method*)

Pada tahap ini data diobservasi, dibagi-bagi variabel mana yang berkorelasi dengan satu sama lainnya. Setelah data diobservasi lalu diimplementasikan model-model yang sesuai dengan jenis data. Pembagian data kedalam data uji coba (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model yang dilakukan dalam poin ini, mengkaitkan data yang satu dengan lainnya, agar menjadi data yang lebih berguna bagi pengguna aplikasi nasabah asuransi.

4. Eksperimendan pengujian model (*Model Testing and Experimen*)

Pada tahap ini model yang diajukan akan diuji untuk melihat hasil berupa aturan yang akan dimanfaatkan dalam pengambilan keputusan. Dalam tahap ini, aplikasi dicoba langsung oleh nasabah asuransi.

5. Evaluasi dan validasi hasil (*Result Evaluation*)

Pada tahap ini dilakukan penilaian terhadap model yang ditetapkan untuk memahami tingkat keakurasian model. Setelah aplikasi dicoba akan muncul kekurangan-kekurangan dari aplikasi tersebut, dan biasa dibenahi. Hasil yang diharapkan dari penelitian ini adalah: jumlah nasabah asuransi aktif, alamat tempat bekerja, dan biodata nasabah asuransi.

Hasil dan Pembahasan

A. Pengumpulan data

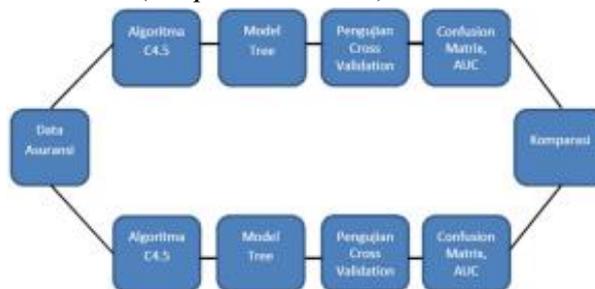
Ada dua tipe dalam pengumpulan data, yaitu pengumpulan data primer dan pengumpulan data sekunder. Data primer adalah data yang dikumpulkan pertama kali untuk melihat apa yang sesungguhnya terjadi sedangkan data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik di terbitkan atau tidak (Martono, 2010). Pengumpulan data primer bisa didapat dari model *interview*

terhadap responden, maupun dengan model observasi terhadap suatu badan yang sedang diteliti (Bustami, 2013). Pada observasi ini digunakan pengumpulan data sekunder.

1. Pengolahan Awal Data

Untuk memperoleh data yang lebih bagus terdapat teknik *preprocessing* yang digunakan pada penelitian ini, yaitu: data *integration and transformation*, untuk menaikkan akurasi dan efisiensi algoritma (Novianto, 2019). Data yang digunakan dalam penulisan ini bernilai kategorikal. Data dirubah kedalam angka menggunakan aplikasi Rapid Miner.

2. Metode Yang diusulkan (*Proposed Method*)



Adapun model algoritma yang dipakai dalam penelitian ini adalah:

- a. Algoritma C4.5 yaitu model untuk merubah data menjadi pohon keputusan dengan aturan-aturannya (*rules*).
- b. *Particle Swarm Optimization* yaitu mendapatkan solusi yang terbaik dengan meningkatkan atribut (*attribute wighting*), berupa partikel yang bergerak kearah posisi yang terbaik dari posisi sebelumnya.

3. Eksperimen Dan Pengujian Model (*Model Testing And Experimen*)

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu mengkalkulasi dan mendapatkan rule-rule yang ada pada model algoritma yang diusulkan. Setelah itu, diuji rule tadi kedalam model *cross validation* untuk mendapatkah hasil yang lebih baik. Evaluasi dan validasi hasil (*Result Evaluation*)

B. Model Algoritma C4.5

Algoritma C4.5 untuk model yang pertama dilakukan. Berikut tahap-tahap yang akan dilakukan sebagai berikut:

- 1. Mengkalkulasi jumlah kasus class ya dan class tidak serta nilai *entropy* dari semua kasus. Kasus dibagi berdasarkan atribut, jumlah kasus 2268 *record*, kelas ya ada 94 *record* dan kelas tidak sebanyak 2174 *record* sehingga didapat *entropy*:

$$\begin{aligned}
 Entropy(i) &= - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j) \\
 &= (-94/2268 \cdot \log_2 (94/2268)) + (-2174/2268 \cdot \log_2 (2174/2268)) \\
 &= 0.2488
 \end{aligned}$$

2. Hitung nilai gain dari setiap atribut, sebagai contoh untuk 3 data nasabah asuransi: $\leq 6989.500 = 2203/2268$ $> 6989.500 = 65/2268$ Atribut nasabah asuransi < 6989.500 terdiri dari 32 class ya dan 2171 class tidak, dan untuk atribut nasabah asuransi > 6989.500 terdiri dari 62 class ya dan 3 class tidak. Nilai Entropynya dapat dikalkulasi sebagai berikut:

$$\begin{aligned}
 Entropy\ split &= \sum_{i=1}^P \frac{n_i}{n} IE(i) \\
 \leq 6989.500 &= \left(\frac{32}{2203} \cdot \log_2 \left(\frac{32}{2203} \right) + \left(\frac{2171}{2203} \right) \cdot \log_2 \left(\frac{2171}{2203} \right) \right) \\
 &= 0.10949 \\
 > 6989.500 &= \left(\frac{62}{65} \cdot \log_2 \left(\frac{62}{65} \right) + \left(\frac{3}{65} \right) \cdot \log_2 \left(\frac{3}{65} \right) \right) \\
 &= 0.26983 \\
 &= \left(\frac{2203}{2268} (0.10949) + \frac{65}{2268} (0.26983) \right) \\
 &= 0.11408 \\
 &= 0.2488 - 0.11408 \\
 &= 0.1347
 \end{aligned}$$

3. Algoritma C4.5 berbasis *Particle Swarm Optimization* Algoritma C4.5 menggunakan metode *particle swarm optimization* untuk model yang kedua dilakukan (JULIA, 2020). Berikut langkah-langkah yang akan dilakukan adalah mengkalkulasi jumlah kasus class ya dan class tidak serta nilai entropy dari semua kasus. kasus dibagi berdasarkan atribut dengan jumlah kasus 2268 record, kelas ya ada 94 record dan kelas tidak sebanyak 2174 record sehingga didapat *entropy*:

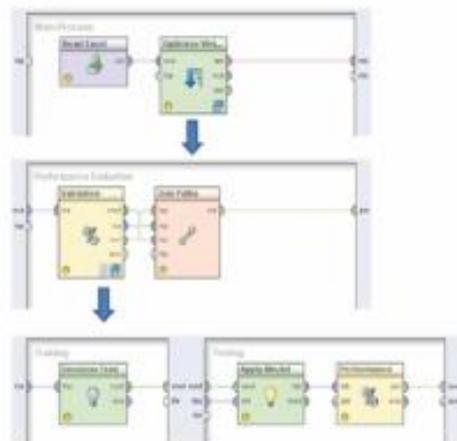
$$\begin{aligned}
 Entropy(i) &= - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j) \\
 &= \left(\frac{94}{2268} \cdot \log_2 \left(\frac{94}{2268} \right) + \left(\frac{2174}{2268} \right) \cdot \log_2 \left(\frac{2174}{2268} \right) \right) \\
 &= 0.2488
 \end{aligned}$$

4. Hitung nilai Gain dari setiap atribut sebagai contoh untuk 3 data nasabah asuransi: $\leq 6989.500 = 2203/2268$ $> 6989.500 = 65/2268$ Atribut nasabah asuransi < 6989.500 terdiri dari 32 class ya dan 2171 class tidak, dan untuk atribut nasabah asuransi > 6989.500 terdiri dari 62 class ya dan 3 class tidak. Nilai entropynya dapat dikalkulasi sebagai berikut:

$$\begin{aligned}
 \text{Entropy split} &= \sum_{i=1}^p \frac{n_i}{n} IE(i) \\
 &\leq 6989.500 = ((-32/2203 \cdot \log_2(32/2203) + (-2171/2203 \\
 &\quad \log_2(2171/2203)) \\
 &= 0.10949 \\
 &> 6989.500 = ((-62/65 \cdot \log_2(62/65) + (-3/65 \cdot \\
 &\quad \log_2(3/65)) \\
 &= 0.26983 \\
 &= ((2203/2268 (0.10949) + (65/2268 (0.26983)) \\
 &= 0.11408 \\
 &= 0.2488 - 0.11408 \\
 &= 0.1347
 \end{aligned}$$

5. Hitung nilai Gain dari masing-masing atribut sebagai contoh untuk nasabah asuransi:

$$\leq 3854.001 = 2149/2268$$



C. Pohon Keputusan C4.5

Adapun tata cara yang didapatkan dari observasi adalah sebagai berikut:

1. IF pekerjaan tetap = lainnya and keluarga mendukung kk = ya then hasil = tidak
2. IF pekerjaan tetap = lainnya and keluarga mendukung kk = tidak kk = selesai sd/mi then hasil = sesuai
3. IF pekerjaan tetap = lainnya and keluarga mendukung kk = ya kk = selesai slta/ma then hasil = tidak

4. IF pekerjaan tetap = lainnya and keluarga mendukung kk = ya kk = selesai sltp/mtsn then hasil = tidak
5. IF pekerjaan tetap = lainnya and keluarga mendukung kk = ya kk = tidak selesai sd/mi then hasil = sesuai
6. IF pekerjaan tetap = ya and perusahaanmengetahui = ya and pendidikan terakhir kk = selesai sd/mi then hasil = sesuai
7. IF pekerjaan tetap = ya and perusahaanmengetahui = tidak and pendidikan terakhir kk = selesai slta/ma then hasil = tidak
8. IF pekerjaan tetap = ya and perusahaanmengetahui = ya and pendidikan terakhir kk = selesai sltp/mtsn then hasil = tidak
9. IF pekerjaan tetap = ya and perusahaanmengetahui = tidak asan and pendidikan terakhir kk = tidak selesai sd/mi then hasil = sesuai
10. IF pekerjaan tetap = ya and perusahaanmengetahui = ya and pendidikan terakhir kk = tidak/belum sekolah then hasil = sesuai
11. IF bekerja minimal 3 tahun = ya then hasil = sesuai
12. IF bekerja minimal 3 tahun = pegawai swasta and aset = tidak then hasil = sesuai
13. IF bekerja minimal 3 tahun = pegawai swasta and aset = ya and fasilitas bab = jamban bersama then hasil = sesuai
14. IF bekerja minimal 3 tahun = ya and aset = ya and fasilitas bab = jamban sendiri then hasil = tidak
15. IF bekerja minimal 3 tahun = tidak and asetmendukung = tidak then hasil = sesuai
16. IF bekerja minimal 3 tahun = ya and asetmendukung = ya and aset = ya and jenis dinding = bambu then hasil = sesuai
17. IF bekerja minimal 3 tahun = tidak and asetmendukung = ya and aset = ya and jenis dinding = tembok and pendidikan terakhir kk = selesai sd/mi and sumber air minum = ledeng/kemasan then hasil = tidak
18. IF bekerja minimal 3 tahun = ya and asetmendukung = ya and aset = ya and jenis dinding = tembok and pendidikan terakhir kk = selesai sd/mi and sumber air minum = sumur terlindung/pompa then hasil = tidak
19. IF bekerja minimal 3 tahun = tidak and asetmendukung = ya and aset = ya and jenis dinding = tembok and pendidikan terakhir kk = selesai slta/ma then hasil = tidak
20. IF bekerja minimal 3 tahun = ya and asetmendukung = ya and aset = ya and jenis dinding = tembok and pendidikan terakhir kk = selesai sltp/mtsn then hasil = tidak
21. IF bekerja minimal 3 tahun = ya and asetmendukung = ya and aset = ya and jenis dinding = tembok and pendidikan terakhir kk = tidak selesai sd/mi then hasil = tidak
22. IF bekerja minimal 3 tahun = tidak then hasil = tidak
23. IF bekerja minimal 3 tahun = ya and pendidikan terakhir kk = selesai sd/mi then hasil = sesuai

D. Evaluasi Confusion Matrix C4.5 Berbasis PSO

Tabel 2
Konversi Confusion Matrix Algoritma Klasifikasi C4.5 Berbasis PSO

	True layak	True tidak	Class precision
Pred SESUAI	293	5	98,88%
Pred TIDAK	12	823	98,83%
Class recall	98,93%	99,15%	

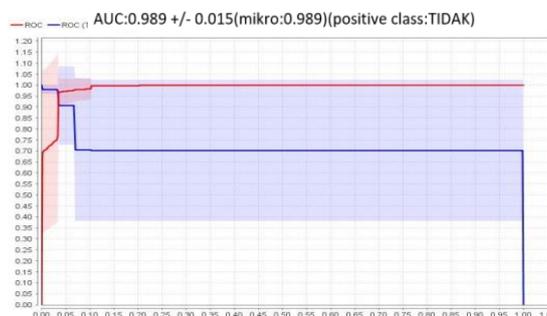
Banyaknya *True Positive* (TP) adalah 293 record dikelompokkan sebagai sesuai dan *False Negative* (FN) sejumlah 9 record dikelompokkan sebagai sesuai tetapi tidak. Berikutnya 823 record untuk *True Negative* (TN) dikelompokkan seperti tidak, dan 3 record *False Positive* (FP) diklasifikasi sebagai tidak ternyata sesuai.

Tabel 3
Hasil Perhitungan Algoritma C4.5 Berbasis PSO

	Nilai(%)
Accuracy	98,93%
Sensitivity	96,95%
Specitivity	99,52%
PPV	97,99%
NPV	98,98%

Melihat Tabel 3 menunjukkan bahwa, tingkat akurasi menggunakan algoritma klasifikasi C4.5 berbasis *particle swarm optimization* adalah sebesar 98,93%.

E. Penilaian dengan ROC C4.5 Berbasis PSO



Gambar 3
Nilai AUC dalam grafik ROC algoritma C4.5 berbasis(PSO)

Pada Gambar 3 menunjukkan grafik ROC dengan nilai AUC (*Area Under Curve*) sejumlah 0,989 dengan golongan diagnosa Excellent Classification.

Kesimpulan

Dari hasil analisa optimasi dapat disimpulkan bahwa nilai akurasi yang didapatkan pada model algoritma C4.5 berbasis PSO adalah 98.93% lebih bagus jika dibandingkan dengan model algoritma C4.5 yaitu 97.84%. Dari hasil tersebut didapatkan perbedaan antara kedua model yaitu senilai 0.4%. Selagi untuk penelaahan menggunakan ROC *curve* bagi kedua model ialah untuk model algoritma C4.5 nilai AUC adalah 0.970 dengan urutan diagnosa *Excellent Classification*, dan untuk model algoritma C4.5 berbasis PSO nilai AUC adalah 0.968 dengan urutan diagnosa *Excellent Classification*. Dari evaluasi ROC *curve* tersebut tampak bahwa model algoritma C4.5 berbasis PSO lebih kecil walaupun tidak merubah tingkat diagnosa, ini disebabkan karena terjadinya *imbalance* terhadap data atau tidak sama antara ya terpilih dan tidak terpilih. lalu diuji kembali menggunakan uji statistik t-test dan mendapatkan hasil bahwa algoritma C4.5 berbasis PSO lebih bagus nilai akurasinya. Akhirnya dapat diambil keputusan akhir bahwa algoritma C4.5 berbasis *particle swarm optimization* lebih akurat dalam menentukan kelayakan asuransi.

BIBLIOGRAFI

- Bustami, B. (2013). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Techsi-Jurnal Teknik Informatika*, 5(2).
- Ermawati, E. (2019). Algoritma Klasifikasi C4. 5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai. *Sistemasi: Jurnal Sistem Informasi*, 8(3), 513–528.
- Indriyanto, J., & Purwanto, C. S. (2014). *Algoritma K-Nearest Neighbor Berbasis Chi-Squared Untuk Prediksi Nasabah Asuransi*. Pascasarjana Teknik Informatika Universitas Dian Nuswantoro.
- Julia, A. (2020). *Analisa History Data Claim Member Asuransi Kesehatan Menggunakan Algoritma C4. 5 Untuk Menentukan Proses Pembayaran Persetujuan Pengobatan*. Universitas Mercu Buana Jakarta.
- Karlia, J. A. (2021). *Penerapan Algoritma C4. 5 Untuk Klasifikasi Keterlambatan Pembayaran Premi Asuransi*. Universitas Katolik Musi Charitas.
- Maralis, R., & Triyono, A. (2019). *Manajemen Resiko*. Deepublish.
- Martono, N. (2010). *Metode Penelitian Kuantitatif: Analisis Isi Dan Analisis Data Sekunder (Sampel Halaman Gratis)*. Rajagrafindo Persada.
- Novianto, R. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menganalisa Bisnis Perusahaan Asuransi. *Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(1), 85–95.
- Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma C4. 5 Sebagai Dasar Pemberian Kredit. *Komputer Teknologi Informasi*, 2(2).
- Santoso, F., Syukur, A., & Fanani, A. Z. (2018). Algoritma C4. 5 Dengan Particle Swarm Optimization Untuk Klasifikasi Lama Menghafal Al-Quran Pada Santri Mahadul Quran. *Jurnal Cyberku*, 14(2), 92–103.
- Siswanti, D. D. (2017). *Penentuan Klasifikasi Pasien Asuransi Dan Non Asuransi Pada Rsud Nganjuk Dengan Metode Algoritma C4. 5*. Sekolah Tinggi Teknik-Pln.
- Sulistiyanto. (2018). Penerapan C4 . 5 Berbasis Particle Swarm Optimization (Pso) Dalam Memprediksi Siswa Lolos Seleksi Perguruan Tinggi. *Semin. Nas. Teknol. Dan Bisnis*, 162–170.
- Umam, K. (2018). *Memahami Dan Memilih Produk Asuransi*. Mediapressindo.
- Utama, Y. H. (2013). *Penerapan Algoritma C4. 5 Untuk Membantu Calon Konsumen Dalam Menentukan Program Asuransi Yang Sesuai (Studi Kasus: Pt. Asuransi Bintang Tbk.)*. University Of Muhammadiyah Malang.

Pengembangan Algoritma C4.5 Berbasis Particle Swarm Optimization untuk Penentuan Kelayakan Asuransi

Utari, D. R., & Wibowo, A. (2020). Pemodelan Prediksi Status Keberlanjutan Polis Asuransi Kendaraan Dengan Teknik Pemilihan Mayoritas Menggunakan Algoritma-Algoritma Klasifikasi Data Mining. *Prosiding Seminar Nasional Teknoka*, 5, 19–24.