

# Prediksi Peminatan Pelanggan dalam Penjualan Produk Sepatu Menggunakan Metode *Decision Tree* Berbasis *Particle Swarm Optimization* pada PT. Baskara Cipta Pratama

Ari Puspita<sup>1</sup>, Jefi<sup>2</sup>, Muhammad Fahmi<sup>3</sup>

**Abstract**—The PSO-based optimization C4.5 model gives a higher value of 78.16% compared to the C4.5 algorithm model that is 73.88. The results obtained differences between the two models by 4.28%. While for evaluation using ROC curve for second model that is, for model of algorithm C4.5 value of AUC is 0,764 with level of diagnosis classification fair, and for model of algorithm C4.5 based on PSO AUC is 0,780 with level of diagnosis of fair classification. It is concluded that ROC curves models shows C4.5 algorithm based on PSO is larger. It can be inferred that C4.5 algorithm based on particle swarm optimization is more accurate in predicting the customers' interest for buying shoes.

**Intisari**— Analisis optimasi model algoritma C4.5 berbasis PSO memberikan nilai akurasi yang lebih tinggi yaitu 78.16% dibandingkan dengan model algoritma C4.5 yaitu 73.88%. Dari hasil tersebut didapatkan selisih antara kedua model yaitu 4,28%. Sementara untuk evaluasi menggunakan ROC curve untuk kedua model yaitu, untuk model algoritma C4.5 nilai AUC adalah 0.764 dengan tingkat diagnosa Fair classification, dan untuk model algoritma C4.5 berbasis PSO nilai AUC adalah 0.780 dengan tingkat diagnosa Fair classification. Dari evaluasi ROC curve tersebut terlihat bahwa model algoritma C4.5 berbasis PSO lebih besar Sehingga dapat disimpulkan bahwa algoritma C4.5 berbasis particle swarm optimization lebih akurat dalam memprediksi minat beli produk sepatu.

**Kata Kunci** — C4.5, Produk, Sepatu PSO

## I. PENDAHULUAN

Dalam meningkatkan penjualan suatu produk harus ditunjang dari sisi produk itu sendiri, Laku atau tidak lakunya suatu produk ditentukan dari segi mutu dan kualitas bahan yang digunakan. Bahan yang tidak baik dapat membuat pemakai produk sepatu tersebut merasa tidak nyaman dan bahkan dapat menyebabkan tidak

lakunya produk sepatu di pasaran, dibandingkan sepatu-sepatu yang dibuat dengan bahan yang bermutu baik, warna yang bervariasi, model yang modern.

Hingga saat ini, ada beberapa kasus yang mempengaruhi pembeli dalam memilih sebuah produk sepatu yang belum diketahui penyebabnya. Ada beberapa faktor dan masalah peminatan pembeli yang dapat memicu pertimbangan dalam memilih sepatu yaitu dari segi merk, warna, dan bahan sepatu.

Perilaku pembelian seseorang dapat dikatakan sesuatu yang unik, karena preferensi dan sikap terhadap obyek setiap orang berbeda[1]. Selain itu konsumen berasal dari berbagai segmen, sehingga apa yang diinginkan dan dibutuhkan juga berbeda. Produsen perlu memahami perilaku konsumen terhadap produk yang ditawarkan dipasaran.

Minat beli adalah tahap kecenderungan responden dalam bertindak sebelum keputusan membeli benar-benar dilakukan[1]. “Minat beli merupakan pernyataan mental dari konsumen yang merefleksikan rencana pembelian sejumlah merek tertentu yang menjadi dasar pemilihan sesuatu, minat membeli menunjukkan pada kecenderungan untuk lebih menyukai produk dengan merek tertentu”, [1]

Pemahaman mengenai bagaimana konsumen melakukan keputusan pembelian merupakan salah satu syarat bagi pemasar untuk meraih keberhasilan dalam melakukan program pemasarannya”[1]. Minat pembelian adalah niat untuk melakukan pembelian pada waktu yang akan datang, Secara khusus, pemasar harus mengidentifikasi siapa yang membuat keputusan pembelian, jenis-jenis keputusan pembelian, dan langkah-langkah dalam proses pembelian. Dengan menggunakan model Decision Tree For Biomaker dari hasil penelitian ini menyebutkan bahwa tingkat akurasi Sensitivity 80% dan akurasi Specificity 54 %.

## II. METODOLOGI PENELITIAN

### A. Metode Pengumpulan Data

Teknik pengumpulan data adalah teknik atau cara- cara yang dapat digunakan untuk menggunakan data Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder[3]. Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder [5]

<sup>1</sup>Jurusan Teknik Informatika, STMIK Antar Bangsa, Kawasan Bisnis CBD Ciledug, Blok A5 No.29-36, Jl. HOS Cokroaminoto Karang Tengah, Tangerang; Telp.021-73453000; e-mail [ari.puspita@gmail.com](mailto:ari.puspita@gmail.com)

<sup>2</sup>Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika, Jl. Kamal Raya No.18, Ringroad Barat, Cengkareng, Jakarta Barate-mail : [jefi.manopo25@gmail.com](mailto:jefi.manopo25@gmail.com)

<sup>3</sup>Program Studi Teknik Informatika, STMIK Nusa Mandiri Jakarta Jl.Kramat Raya B No.18 Jakarta Pusat; e-mail : [fahmy2205@gmail.com](mailto:fahmy2205@gmail.com)

Data yang diperoleh merupakan data primer karena data tersebut diambil dari sebuah perusahaan yang terdapat di daerah Purwakarta. Atribut yang terdapat didalamnya merupakan faktor-faktor penyebab laku atau tidak lakunya produk sepatu. Faktor penyebab laku atau tidak lakunya produk sepatu biasanya didominasi dari beberapa atribut, salah satunya warna, kualitas bahan, model, dan *brand image* produk itu sendiri. Faktor yang sering terjadi adalah model yang kurang sesuai keinginan. Dan faktor lain pun dapat mempengaruhi terjadinya laku atau tidak lakunya produk yaitu mempunyai *brand image* yang baik, bahan yang kuat, kelengkapan warna produk dan ukuran produk.

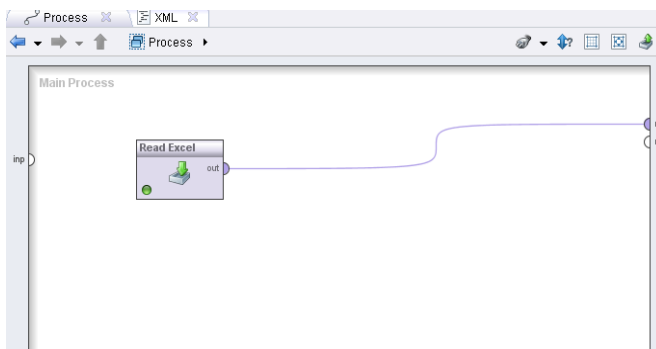
Untuk lebih jelasnya dapat dilihat daftar data penjualan produk sepatu dalam Tabel 1.

### III. HASIL DAN PEMBAHASAN

#### A. Pengolahan Data Awal

Ditahapan ini dijelaskan tentang tahapan awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pengelompokan dan penentuan atribut data.

1. Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 490 record, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut,[2]



Gbr 1. Proses *Missing value* dengan Rapid Miner

Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Data penjualan yang ada akan diidentifikasi apakah ada data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*), lihat pada gambar 2

Gbr 2. Hasil Pengecekan *Missing Value* dengan Rapid Miner

2. Data *integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam *software Rapidminer*. Tabel kategorikal atribut terlihat pada tabel 2
3. Data *size reduction and discretization*, untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Pada tahap ini data yang masih berupa nilai numerik dan juga *continue* ditransformasikan kedalam bentuk kategorikal dan dibuat skala atau interval sehingga menghasilkan *range* yang lebih kecil sebagai bahan pembelajaran algoritma C4.5.

TABEL 2. ATRIBUT YANG DIGUNAKAN

No	Atribut	Nilai
1	Model	Casual, Flats shoes, Loafers, Marry_Jane, Wedges
2.	Warna	Biru, Merah_Muda, Putih, Abu-Abu, Coklat, Hitam, Mocca, Merah, kuning_tua, hijau, Coklat, Hitam, Salem, Biru, Tosca, Super_Hitam, Merah_Hati, Biru_Tua
3.	Size	36, 37, 38, 39, 40, 41, 42, 43, 44
4.	Kelamin	Wanita dan Pria
5.	Bahan Upper	Canvas_8_Oz, Jevelin, Lopez, Kentucky, New_Nasa, Jubille, Tomcat, Jubille
6.	Merk	Nevada, ST. Yves, Connexion
7.	Kategori	Formal dan Tidak Formal
8.	Hasil	Laku atau Tidak laku

Pengolahan awal data berupa pembentukan sumber data *random (set the random seed)* dan pembentukan variabel pemilihan (*partition variabel*). Untuk mendapatkan data yang dapat tersusun dengan baik maka diperlukan pembentukan sumber data random karena sudah mengalami pengacakan data secara statistik. Data yang digunakan berjumlah 490 record, dibagi menjadi dua untuk data *training* (80%) dan data *testing* (20%), sebanyak 490 record akan digunakan untuk data training sebanyak 80% atau 392 record dan 20% atau 98 record digunakan sebagai data testing. Pengambilan data untuk data testing menggunakan teknik *Systematic Random Sampling*.

TABLE I  
ATRIBUT DAN PENJUALAN PRODUK SEPATU  
PADA PT. BASKARA

NO	JENIS	MODEL	WARNA	SIZE	KELAMIN	BAHAN	MERK	KATEGORI	HASIL
1	Flats oes	Pita	Navy	36	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
2	Flats oes	Pita	Navy	37	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
3	Flats oes	Pita	Navy	38	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
4	Flats oes	Pita	Navy	39	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
5	Flats oes	Pita	Navy	40	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
6	Flats oes	Pita	Navy	41	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
7	Flats oes	Pita	Pink	36	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
8	Flats oes	Pita	Pink	37	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
9	Flats oes	Pita	Pink	38	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
10	Flats oes	Pita	Pink	39	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
11	Flats oes	Pita	Pink	40	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
12	Flats oes	Pita	Pink	41	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
13	Flats oes	Pita	Putih	36	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
14	Flats oes	Pita	Putih	37	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku
15	Flats oes	Pita	Putih	38	Wanita	Canvas 8 Oz	St. Yves	Formal	Laku
16	Flats oes	Pita	Putih	39	Wanita	Canvas 8 Oz	St. Yves	Formal	Tidak Laku

Sumber Hasil Penelitian (2017)

Teknik *systematic random sampling* merupakan modifikasi dari teknik random sampling, dengan cara memilih subjek dari daftar populasi secara sistematis bukan secara acak. Pemilihan secara acak hanya digunakan untuk memilih data pertama saja. Penentuan data berikutnya dengan cara memanfaatkan interval sampel, yaitu angka yang menunjukkan jarak antara nomor-nomor urut yang terdapat dalam kerangka sampling yang dijadikan sebagai patokan dalam menentukan data kedua dan seterusnya hingga data ke-n[4]

Untuk menentukan interval sampel menggunakan rumus berikut:

$$f = N/sn \quad (1)$$

dimana:

f = frekuensi interval

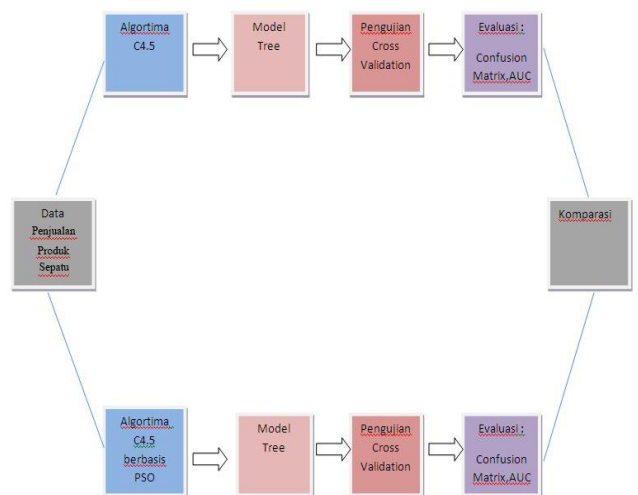
N = Jumlah total populasi

sn = Jumlah sample yang diperlukan

Total data testing yang diperlukan sebanyak 20% dari 490 record yaitu 98 record, dengan menggunakan rumus 1 untuk mendapatkan interval maka dari 490 record dibagi dengan 98 record sehingga didapat intervalnya adalah 5. Sehingga untuk pemilihan data berikutnya didasarkan pada nomor kelipatan lima. Tabel.3 menunjukkan sample data penjualan produk sepatu pada PT. Baskara yang telah melalui tahapan data preparation. Data penjualan produk sepatu secara lengkap dapat dilihat pada lampiran.

### B. Metode Yang Diusulkan

Pada tahap modeling ini dilakukan pemrosesan data training sehingga akan membahas metode algoritma yang diuji dengan memasukan data penjualan produk sepatu kemudian di analisa dan dikomparasi. Berikut ini bentuk gambaran metode algoritma yang akan diuji



Gbr 3 Metode Algoritma Yang Diusulkan

TABEL 3. SAMPEL DATA TRAINING

No	Model	Warna	Size	Kelamin	Bahan Upper	Merk	Kategori	Hasil
1	Flatshoes	Biru	36	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
2	Flatshoes	Biru	37	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
3	Flatshoes	Biru	38	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
4	Flatshoes	Biru	39	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
5	Flatshoes	Biru	40	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
6	Flatshoes	Biru	41	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
7	Flatshoes	Merah_Muda	36	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
8	Flatshoes	Merah_Muda	37	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
9	Flatshoes	Merah_Muda	38	Wanita	Canvas_8_Oz	St.Yves	Formal	Tidak Laku
10	Flatshoes	Merah_Muda	39	Wanita	Canvas_8_Oz	St.Yves	Formal	Tidak Laku
11	Flatshoes	Merah_Muda	40	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
12	Flatshoes	Merah_Muda	41	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
13	Flatshoes	Putih	36	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
14	Flatshoes	Putih	37	Wanita	Canvas_8_Oz	St.Yves	Formal	Tidak Laku
15	Flatshoes	Putih	38	Wanita	Canvas_8_Oz	St.Yves	Formal	Laku
16	Flatshoes	Putih	39	Wanita	Canvas_8_Oz	St.Yves	Formal	Tidak Laku
17	Flatshoes	Putih	40	Wanita	Canvas_8_Oz	St.Yves	Formal	Tidak Laku

C. Eksperimen dan Pengujian Metode

Metode penelitian eksperimen dapat diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan. Berdasarkan pendapat tersebut, dapat dipahami bahwa penelitian eksperimen selalu dilakukan dengan memberikan perlakuan terhadap subyek penelitian kemudian melihat pengaruh dari perlakuan tersebut. [2]

Tahap modeling untuk menyelesaikan prediksi laku atau tidak lakunya produk sepatu dengan menggunakan model *Decision Tree* berbasis *Particle Swram Optimization* (PSO)

1. *Decision Tree* adalah “salah satu teknik tertua dan yang paling tertua untuk model learning discriminatory, yang mana model tersebut telah dikembangkan secara mandiri di statistik dan di komunitas machine learning”. metode yang menghitung *entropy* dan gain *entropy* yang kemudian nilai gain tertinggi akan membentuk sebagai akar dari node yang kemudian akan berbentuk menjadi sebuah pohon keputusan. *Decision Tree* merupakan algoritma pengklasifikasian yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan [2]
2. *Particle Swarm Optimimization* (PSO) adalah algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel [2]

D. Evaluasi dan validasi hasil

Model yang diusulkan pada penelitian tentang prediksi penjualan produk sepatu adalah model *Decision Tree*. Data diolah dengan menggunakan metode *Decision Tree*, setelah diolah dan menghasilkan model. Maka untuk meningkatkan akurasi digunakan metode *Particle Swarm Optimimization* (PSO). model yang dihasilkan tersebut dilakukan pengujian menggunakan *k-fold cross validation*, kemudian dilakukan

evaluasi dan validasi hasil dengan confusion matrix dan kurva ROC. Tahap selanjutnya adalah memperbandingkan hasil akurasi dan AUC, sehingga diperoleh model dari metode klasifikasi yang mana yang memperoleh nilai akurasi dan AUC tertinggi.

E. Eksperimen dan Pengujian Model

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan rule-rule yang ada pada model algoritma yang diusulkan, Data eksperimen yang uji yaitu data penjualan produk sepatu dari PT. Baskara Cipta Pratama, Setelah diuji rule tadi kedalam model *cross validation* untuk mendapatkah hasil yang lebih baik.

$$E_{abu-abu} [6,6] = (-6/12 \log_2 6/12) + (-6/12 \log_2 6/12) = 1$$

$$E_{biru} [92,21] = (-92/113 \log_2 92/113) + (-21/113 \log_2 21/113) = 0,6926$$

$$E_{biru tua} [2,5] = (-2/7 \log_2 2/7) + (-5/7 \log_2 5/7) = 0,8631$$

$$E_{biru tosca} [14,4] = (-14/18 \log_2 14/18) + (-4/18 \log_2 4/18) = 0,7642$$

$$E_{coklat} [31,11] = (-31/42 \log_2 31/42) + (-11/42 \log_2 11/42) = 0,8296$$

$$E_{hijau} [9,9] = (-9/18 \log_2 9/18) + (-9/18 \log_2 9/18) = 1$$

$$E_{hitam} [50,12] = (-50/62 \log_2 50/62) + (-12/62 \log_2 12/62) = 0,7088$$

$$E_{kuning tua} [7,17] = (-7/24 \log_2 7/24) + (-17/24 \log_2 17/24) = 0,8708$$

$$\text{Emerah hati } [15,9] = (-15/24 \log_2 15/24) + (-9/24 \log_2 9/24) \\ = 0,9544$$

$$\text{Emerah } [43,23] = (-43/66 \log_2 43/66) + (-23/66 \log_2 23/66) \\ = 0,9327$$

$$\text{Emerah muda } [8,4] = (-8/12 \log_2 8/12) + (-4/12 \log_2 4/12) \\ = 0,9182$$

$$\text{Emocca } [10,2] = (-10/12 \log_2 10/12) + (-2/12 \log_2 2/12) \\ = 0,6500$$

$$\text{Eputih } [6,13] = (-6/19 \log_2 6/19) + (-13/19 \log_2 13/19) \\ = 0,8997$$

$$\text{Esalem } [3,15] = (-3/18 \log_2 3/18) + (-15/18 \log_2 15/18) \\ = 0,6500$$

$$\text{Esuper hitam } [3,15] = (-3/18 \log_2 3/18) + (-15/18 \log_2 15/18) \\ = 0,6500$$

$$\text{Etosca } [12,13] = (-12/25 \log_2 12/25) + (-13/25 \log_2 13/25) \\ = 0,9988$$

Nilai *Entropy* nya dapat dihitung sebagai berikut:

$$\text{Entropy split} = \sum_{i=1}^p \frac{n_i}{n} IE(i) \quad (2)$$

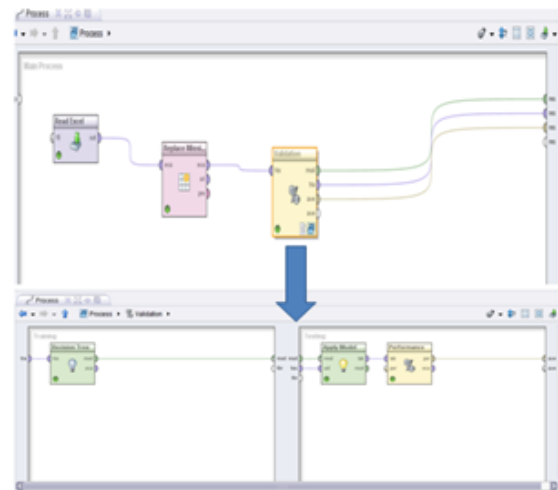
*E split* warna

$$= ((12/490 (1) + (113/490 (0,6926) + (7/490 (0,8631) + \\ (18/490 (0,7642) + (42/490 (0,8296) + (18/490 (1) + \\ (62/490 (0,7088) + (24/490 (0,8708) + (24/490 (0,9544) + \\ (66/490 (0,9327) + (12/490 (0,9182) + (12/490 (0,6500) + \\ (19/490 (0,8997) + (18/490 (0,6500) + (18/490 (0,6500) + \\ (25/490 (0,9988)) = 0,8091$$

$$\text{Gain warna} = 0,9469 - 0,8091 \\ = 0,1378$$

Setelah diolah maka dilakukan teknik pengujian dengan metode *k-fold cross validation* pada tools *RapidMiner*, pengolahan pengujian untuk metode algoritma C4.5 terlihat seperti gambar 4

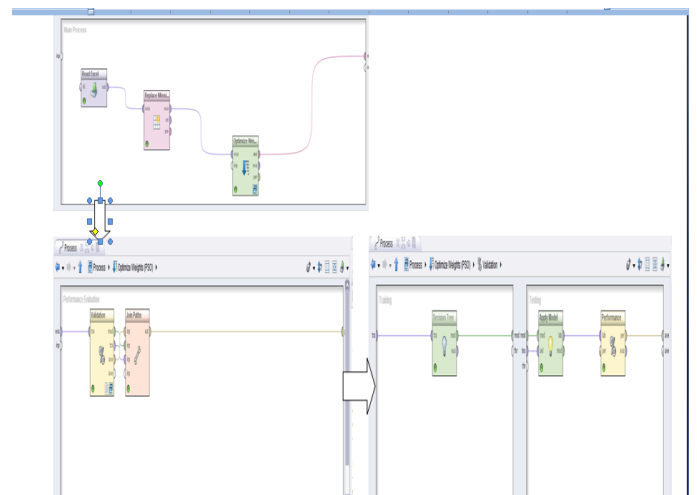
Langkah awal pengolahannya adalah dengan mengambil data training yang akan digunakan pada modul *Read Excel* kemudian dihubungkan dengan modul *validation*, modul *validation* didalamnya berisi metode yang digunakan dalam pengujian data.



Gbr 4 Pengujian *K-Fold Cross Validation* Metode Algoritma C4.5

#### F. Algoritma C4.5 berbasis *Particle Swarm Optimization*

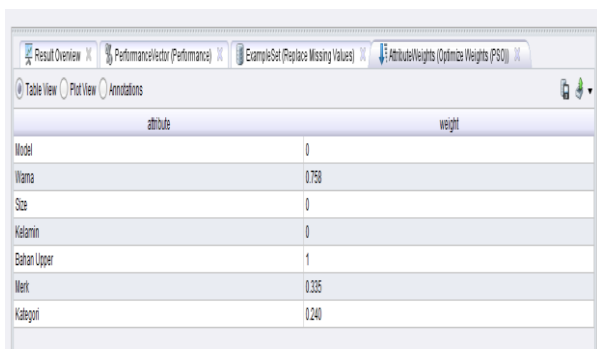
Pada penelitian prediksi penjualan produk sepatu menggunakan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO). *Particle Swarm Optimization* (PSO), memiliki pencarian kinerja yang lebih baik untuk memecahkan banyak masalah optimasi dengan lebih cepat dan tingkat konvergensi yang stabil.



Gbr 5 Desain model algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO)

Berdasarkan hasil dari uji coba dengan menggunakan algoritma C4.5 maka data *training* tersebut akan diseleksi dengan attribute yang digunakan yaitu model sepatu, warna, *size*, bahan upper, kelamin, merk, kategori, dan 1 atribut sebagai label yaitu hasil.

TABEL 4.  
BOBOT ATRIBUT DENGAN METODE C4.5 DENGAN PSO



Model	atribute	weight
Model		0
Warna		0.758
Size		0
Kelamin		0
Bahan Upper		1
Merk		0.335
Kategori		0.240

Dari hasil uji coba menggunakan algoritma C4.5 berbasis *particle swarm optimization* diperoleh beberapa atribut -atribut yang berpengaruh terhadap bobot atribut yaitu: warna, bahan upper, merk, kategori. Berdasarkan hasil uji coba model, size, dan kelamin tidak berpengaruh terhadap bobot attribute.

G. Evaluasi dan Validasi Hasil

Pengujian model dengan C4.5 dan C4.5 berbasis *Particle Swarm Optimization* (PSO) dengan menentukan nilai *accuracy* dan AUC, untuk menghasilkan prediksi penjualan produk sepatu. Dalam menentukan tingkat keakurasian dalam model C4.5 dan algoritma C4.5 berbasis *particle swarm optimization* (PSO). Berikut adalah desain model Metode pengujiannya menggunakan *cross validation*.

H. Hasil Pengujian Model Algoritma C4.5

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*).

1. Confusion Matrix

Tabel 5 adalah perhitungan data *training* pada Table 4 dengan jumlah data 490 record. Berikut Tabel yang didapat:

accuracy: 73.88% +/- 5.31% (mikro: 73.88%)

	true Laku	true Tidak Laku	class precision
pred Laku	256	73	77.81%
pred Tidak Laku	55	106	65.84%
class recall	82.32%	59.22%	

TABEL 5  
MODEL CONFUSION MATRIX UNTUK ALGORITMA C4.5

Jumlah *True Positive* (TP) adalah 256 record diklasifikasikan sebagai Laku dan *False Negative* (FN) sebanyak 73 record diklasifikasikan sebagai Tidak laku tetapi laku. Berikutnya 106 record untuk *True Negative* (TN) diklasifikasikan sebagai tidak laku, dan 55 record *False Positive* (FP) diklasifikasikan sebagai laku ternyata laku. Berdasarkan Tabel 5. tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 73.88%, dan dapat dihitung untuk mencari nilai

*accuracy, sensitivity, specificity, ppv, dan npv*. Hasilnya dapat dilihat pada table 6

TABEL 6.  
NILAI ACCURACY, SENSITIVITY, SPECIFICITY, PPV, DAN NPV

	%
Accuracy	73.88
Sensitivity	77.81
Specificity	65.84
PPV	82.32
Npv	59.22

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+ FN} \\ &= \frac{256 + 106}{256+106+55+73} = \frac{362}{490} = 0,7387 \\ &= 0,7387 \times 100\% = 73,88\% \end{aligned}$$

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP + FN} \\ &= \frac{256}{256+73} = \frac{256}{329} = 0,7781 \\ &= 0,7781 \times 100\% = 77,81\% \end{aligned}$$

$$\begin{aligned} \text{Specificity} &= \frac{TN}{TN + FP} \\ &= \frac{106}{106+55} = \frac{106}{161} = 0,6584 \\ &= 0,6584 \times 100\% = 65,84\% \end{aligned}$$

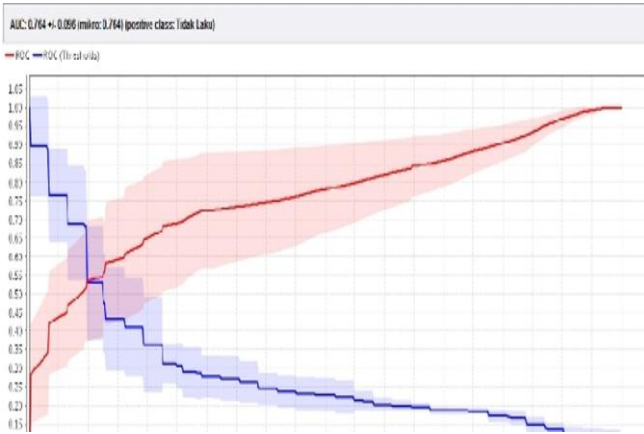
$$\begin{aligned} \text{PPV} &= \frac{TP}{TP + FP} \\ &= \frac{256}{256+55} = \frac{256}{311} = 0,8232 \\ &= 0,8232 \times 100\% = 82,32\% \end{aligned}$$

$$\begin{aligned} \text{NPV} &= \frac{TN}{TN + FN} \\ &= \frac{106}{106+73} = \frac{106}{179} = 0,5922 \\ &= 0,5922 \times 100\% = 59,22\% \end{aligned}$$

2. Evaluasi ROC Curve  
 Dari Gambar 5 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.780. Tingkat akurasi dapat di diagnosa sebagai berikut:  
 Akurasi 0.90 – 1.00 = *Excellent classification*  
 Akurasi 0.80 – 0.90 = *Good classification*

Akurasi 0.70 – 0.80 = *Fair classification*  
 Akurasi 0.60 – 0.70 = *Poor classification*  
 Akurasi 0.50 – 0.60 = *Failure*

Dari gambar 6 terdapat gambar grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.764 dengan diagnosa hasilnya *Fair classification*



Gbr 6 Nilai AUC Dalam Grafik ROC Algoritma C4.5

I. Algoritma C4.5 berbasis *Particle Swarm Optimization*

Pada penelitian prediksi penjualan produk sepatu menggunakan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO). *Particle Swarm Optimization* (PSO), memiliki pencarian kinerja yang lebih baik untuk memecahkan banyak masalah optimasi dengan lebih cepat dan tingkat konvergensi yang stabil.



Gambar 7 Desain model algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO)

Berdasarkan hasil dari uji coba dengan menggunakan algoritma C4.5 maka data *training* tersebut akan diseleksi dengan attribute yang digunakan yaitu model sepatu, warna, *size*, bahan upper, kelamin, merk, kategori, dan 1 *atribut* sebagai label yaitu hasil.

TABEL 7  
BOBOT *ATRIBUT* DENGAN METODE C4.5 DENGAN PSO

	BOBOT	BOBOT
Warna	1	
Ukuran	1.75	
Ukuran	1	
Merk	1	
Bahan Upper	1	
Kelamin	0.02	
Kategori	0.04	

Dari hasil uji coba menggunakan algoritma C4.5 berbasis *particle swarm optimization* diperoleh beberapa *atribut -atribut* yang berpengaruh terhadap bobot *atribut* yaitu: warna, bahan upper, merk, kategori. Berdasarkan hasil uji coba model, *size*, dan kelamin tidak berpengaruh terhadap bobot attribute.

J. Evaluasi dan Validasi Hasil

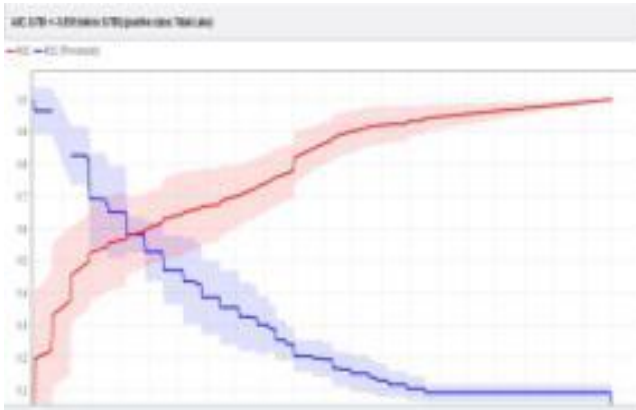
Pengujian model dengan C4.5 dan C4.5 berbasis *Particle Swarm Optimization* (PSO) dengan menentukan nilai *accuracy* dan AUC, untuk menghasilkan prediksi penjualan produk sepatu. Dalam menentukan tingkat keakurasian dalam model C4.5 dan algoritma C4.5 berbasis *particle swarm optimization* (PSO). Berikut adalah desain model Metode pengujiannya menggunakan *cross validation*

K. Analisis Evaluasi dan Validasi Model

Dari hasil pengujian diatas, baik evaluasi menggunakan *confusion matrix* maupun *ROC curve* terbukti bahwa hasil pengujian algoritma C4.5 berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma C4.5. Nilai akurasi untuk model algoritma C4.5 sebesar 73.88% dan nilai akurasi untuk model algoritma C4.5 berbasis PSO sebesar 78,16% dengan selisih akurasi 4.28%, dapat dilihat pada Tabel 8 dibawah ini:

TABEL 8  
PENGUJIAN ALGORITMA C4.5 DAN C4.5 BERBASIS PSO

	<i>Accuracy</i>	AUC
<b>C4.5</b>	73.88%	0.764
<b>C4.5 berbasis PSO</b>	78.16%	0.780



Gambar 8. ROC curve (Algoritma C4.5 dan Algoritma C4.5 berbasis PSO)

Untuk evaluasi menggunakan ROC curve sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma C4.5 menghasilkan nilai 0.764 dengan nilai *Fair classification*, sedangkan untuk algoritma C4.5 berbasis PSO (*Particle Swarm Optimization*) menghasilkan nilai 0.780 dengan nilai *Fair classification*, dan selisih nilai keduanya sebesar 0.016.

#### IV. KESIMPULAN

Dari hasil analisis optimasi model algoritma C4.5 berbasis PSO memberikan nilai akurasi yang lebih tinggi yaitu 78.16% dibandingkan dengan model algoritma C4.5 yaitu 73.88%. Dari hasil tersebut didapatkan selisih antara kedua model yaitu 4,28%. Sementara untuk evaluasi menggunakan ROC curve untuk kedua model yaitu, untuk model algoritma C4.5 nilai AUC adalah 0.764 dengan tingkat diagnosa *Fair classification*, dan untuk model algoritma C4.5 berbasis PSO nilai AUC adalah 0.780 dengan tingkat diagnosa *Fair classification*. Dari evaluasi ROC curve tersebut terlihat bahwa model algoritma C4.5 berbasis PSO lebih besar. Sehingga dapat disimpulkan bahwa algoritma C4.5 berbasis *particle swarm optimization* lebih akurat dalam memprediksi minat beli prodak sepatu.

#### REFERENSI

- [1] Arifin. 2015. *Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Teksstur*: Jurnal Informatika. Vol.II. No.2.
- [2] Fitrianiingsih dan Musdalifah. 2015. *Efektivitas Penggunaan Media Video Pada Pembelajaran Pembuatan Strapless Siswa Kelas XII SMK Negeri 1 Jambu*: Jurnal UNNES. FFEJ 4 (1) (2015).
- [3] Puspita dan Wahyudi. 2015. *Algoritma C4.5 Berbasis Decision Tree Untuk Prediksi Kelahiran Bayi Prematur*: Konferensi Nasional Ilmu Pengetahuan dan Teknologi 2015
- [4] Santi dan Yulianto. 2014. *Sistem Informasi Penjualan Alat-Alatperikanan Pada Usaha Dagang (UD)Jaring Mas Pacitan*: Speed Journal–Sentra Penelitian Engineering dan Edukasi. Vol.11. No.1.



Ari Puspita, M.Kom. Tahun 2011 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri. Tahun 2015 Lulus dari Magister Ilmu Komputer Program Pascasarjana STMIK Nusa Mandiri.



Jefi, M.Kom . Tahun 2014 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta. Tahun 2018 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Saat ini mengajar di Universitas Bina Sarana Informatika Kampus Bekasi.



Muhammad Fahmi M.Kom Tahun 2011 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta. Tahun 2014 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Saat ini mengajar di STMIK Nusa Mandiri Jakarta.