

## Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah

Hendri, M.Kom<sup>1)</sup>, Yana Iqbal Maulana, M.Kom<sup>2)</sup>  
 STMIK Nusa Mandiri Jakarta<sup>1)</sup> Universitas Bina Sarana Informatika<sup>2)</sup>  
[hendri.hed@nusamandiri.ac.id](mailto:hendri.hed@nusamandiri.ac.id)<sup>1)</sup> [yana.yim@bsi.ac.id](mailto:yana.yim@bsi.ac.id)<sup>2)</sup>

**Abstract** - Until now, there have been several cases that have affected buyers in choosing a shoe product whose cause is unknown. There are several factors and problems of buyer interest that can trigger consideration in choosing shoes, namely in terms of brand, color, and shoe material.

An understanding of how consumers make purchasing decisions is one of the requirements for marketers to achieve success in conducting their marketing programs. "The interest in purchasing is the intention to make a purchase in the future, In particular, marketers must identify who made the purchasing decision, the types of purchasing decisions, and the steps in the buying process. The optimization analysis of the C4.5 algorithm model gives an accuracy value with the C4.5 algorithm model which is 73.88%. From evaluation using the ROC curve for the model, that is, for the C4.5 algorithm model, the AUC value is 0.764 with Fair classification diagnosis. From the evaluation of the ROC curve it can be seen that the C4.5 algorithm model can be concluded that the C4.5 algorithm is quite accurate in predicting buying interest in shoe products

**Keywords:** ROC, C4.5, Shoe Products

**Abstrak** – 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.

Pemahaman mengenai bagaimana konsumen melakukan keputusan pembelian merupakan salah satu syarat bagi pemasar untuk meraih keberhasilan dalam melakukan program pemasarannya". 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. Analisis optimasi model algoritma C4.5 memberikan nilai akurasi dengan model algoritma C4.5 yaitu 73.88%. Dari evaluasi menggunakan ROC curve untuk model yaitu, untuk model algoritma C4.5 nilai AUC adalah 0.764 dengan tingkat diagnosa Fair classification. Dari evaluasi ROC curve tersebut terlihat bahwa model algoritma C4.5 dapat disimpulkan bahwa algoritma C4.5 cukup akurat dalam memprediksi minat beli produk sepatu

**Kata kunci:** ROC, C4.5, Produk Sepatu

### 1.a Latar Belakang

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. 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.

Pemahaman mengenai bagaimana konsumen melakukan keputusan pembelian merupakan salah satu syarat bagi pemasar untuk meraih keberhasilan dalam melakukan program pemasarannya". 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 %.

### 1.b Rumusan Masalah

Berdasarkan hasil penelitian yang sudah dilakukan oleh beberapa peneliti terlihat bahwa model *decision tree* sudah termasuk kedalam kategori akurasi yang cukup baik tetapi ada juga yang lebih baik yaitu *logistic regression*. Sedangkan pertanyaan penelitian adalah seberapa akurat model algoritma C4.5.

pada penelitian sebelumnya yang berjudul "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE" (Fandy Ferdian Harryanto, Seng Hansun, 2017). Dalam penelitian ini di jelaskan bahwa Implementasi algoritma C4.5 untuk melakukan prediksi terhadap calon pegawai baru pada PT WISE telah berhasil dilakukan. Hasil tingkat keberhasilan prediksi calon pegawai baru di PT WISE secara keseluruhan yang telah diukur menggunakan metode ten-fold cross validation adalah sebesar 71%.

### 1.c Batasan Masalah

Dalam penelitian ini dibatasi pada algoritma C4.5 dengan cara menganalisis sejumlah atribut yang menjadi parameter Untuk Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah, diantaranya: Model, Warna, Kelamin, Size, Bahan, Merk, dan Kategori. dengan cara menganalisis sejumlah atribut yang menjadi bobot atribut (attribute weight) untuk Prediksi Prediksi Pengangkatan Karyawan Kontrak Menjadi Karyawan Tetap.

### 1.d. Tujuan

Tujuan dari penelitian ini adalah menerapkan seleksi atribut yang optimal dan membobot atribut dari data set pada metode algoritma C4.5 guna meningkatkan akurasi Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah.

### 1.e. Manfaat Penelitian

Manfaat praktis dari hasil penelitian ini adalah dapat digunakan oleh analis perusahaan untuk melakukan analisa Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah.

### 1. f. Metode Penelitian

Penggunaan data mining untuk menentukan Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah menggunakan *Decision Tree*. Peneliti lain yang sebelumnya yaitu tentang "Perancangan Sistem Pendukung Keputusan Menentukan Status Karyawan Kontrak jadi Karyawan Tetap Dengan Metode Simple Additive Weighting (SAW)" (Rudi Hartoyo, 2013). Dalam penelitian ini di jelaskan bahwa bagaimana Sistem pendukung keputusan ini dirancang menggunakan metode Simple Additive Weighting

(SAW) untuk menghitung hasil dalam menentukan karyawan kontrak SPG menjadi karyawan tetap.

### 2.a. Dasar Teori

Beberapa literatur yang mendasari penelitian ini akan dibahas pada bab ini, meliputi sistem pendukung keputusan, pohon keputusan, algoritma C4.5.

### 2.b. Sistem Pendukung Keputusan

Sistem pendukung keputusan adalah sebuah sistem yang dapat membantu manusia untuk mengambil keputusan secara objektif. Konsep sistem seperti ini pertama kali dicetuskan pada tahun 1970-an oleh Michael S.Scott Morton, Michael pertama kali menyebut sistem seperti ini dengan nama *Management Decision System*. Maksud dan tujuan dari adanya "sistem pengambil keputusan adalah mendukung pengambil keputusan untuk memilih alternatif keputusan menggunakan model-model pengambilan keputusan dan untuk menyelesaikan masalah yang bersifat terstruktur, semi terstruktur, maupun tidak terstruktur" (Sulistiyani, T. dan Ambar, R., 2003).

### 2.c. Data mining

Lumbantoruan (2015:67) menyebutkan bahwa *data mining* sering disebut *Knowledge Discovery in Database* (KDD), yaitu kegiatan yang meliputi pengumpulan, pemakaian data histori untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar.

*Data mining* merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. *Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya, Lumbantoruan (2015:67).

Terdapat empat pelompokan dalam data mining yaitu klasifikasi, asosiasi, *clustering* dan prediksi. Defiyanti, (2017:257):

#### 1. Klasifikasi

Proses klasifikasi didasarkan: Kelas-variabel dependen dari model-yang merupakan variabel kategori mewakili yang 'label' memakai objek setelah klasifikasi, contohnya loyalitas pelanggan, kelas bintang (galaksi), kelas gempa bumi, Defiyanti, (2017:258).

Menganalisa konsumen yang akan membeli komputer atau menganalisa apakah seseorang mengidap kanker merupakan contoh proses klasifikasi dalam *data mining* dimana klasifikasi

dapat menghasilkan sebuah prediksi seperti “beresiko” dan “tidak beresiko”, atau “ya” dan “tidak”.

## 2. Asosiasi

Setiap asosiasi antara fitur-fitur yang dicari, bukan hanya satu yang memprediksi nilai kelas tertentu. Pada prinsipnya, penemuan aturan asosiasi/asosiasi mempelajari aturan bagaimana kita memahami proses mengidentifikasi aturan antara ketergantungan yang berbeda dari fenomena kelompok. Dengan demikian, mari kita perkirakan kumpulan set yang kita punya masing-masing berisi sejumlah objek/benda-benda. Jadi tujuan kita untuk mencari peraturan yang menghubungkan (asosiasi), obyek ini berdasarkan peraturan ini, untuk dapat memprediksi terjadinya objek/item, berdasarkan kejadian lain, Defiyanti, (2017:257).

## 3. Clustering

Cluster adalah menemukan kelompok (kelompok) objek, berdasarkan kemiripan (semacam kemiripan), sehingga dalam setiap kelompok ada kemiripan yang besar, sementara kelompok cukup berbeda dari satu sama lain, Defiyanti, (2017:257).

## 4. Prediksi

Prediksi/perkiraan model yang berkaitan dengan kemampuan untuk memprediksi tanggapan terbaik (*output*), yang paling dekat ke kenyataan, berdasarkan input data.

### 2.d. Algoritma C4.5

Pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan bekerja mulai dari akar paling atas, jika diberikan sejumlah data uji, misalnya X dimana kelas dari data X belum diketahui, maka pohon keputusan akan menelusuri mulai dari akar sampai node dan setiap nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan pohon keputusan, kemudian pohon keputusan akan memprediksi kelas dari tupel X.

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu

inputan disebut *node internal* atau *test node*, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu *class*, Maimon dan Rockach dalam Arifin, (2015). Pohon keputusan dibuat dengan membagi nilai-nilai atribut menjadi cabang untuk setiap kemungkinan. Cara kerja pohon keputusan yaitu dengan melakukan penelusuran dari akar hingga ke cabang sampai *class* suatu objek ditemukan, *instance* diklasifikasikan dengan mengarahkan dari akar pohon sampai ke daun sesuai dengan hasil tes melalui *node internal*.

Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Iriadi Nandang dan Nuraeni, 2016:133) yaitu:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j) \quad \text{Error! Reference source not found.} \quad (2.1)$$

Keterangan:

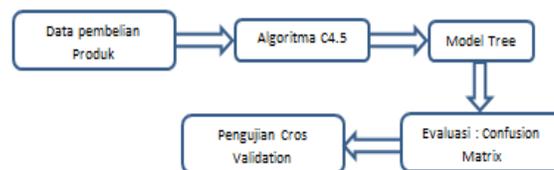
i = himpunan kasus

m = jumlah partisi i

f(i,j) = proposi j terhadap i

### 3. Metode Penelitian

Metode Penelitian yang digunakan dalam penerapan algoritma C4.5 untuk Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah menggunakan rancangan penelitian yang ditunjukkan pada Gambar 1 berikut :



Gambar 1. Metode Algoritma yang diusulkan

#### 4. Implementasi dan Hasil

Implementasi dan hasil yang didapat dalam algoritma klasifikasi C4.5 adalah sebagai berikut :

##### 4.a. Pengumpulan Data

Pengumpulan data dilakukan dengan cara meminta data hasil penilaian karyawan kontrak pada PT. Zoelindo Elite Anugerah dengan variabel Model, Warna, Kelamin, Size, Bahan, Merk, dan Kategori. Jumlah data yang akan digunakan sebagai sampel adalah 70 data.

**Tabel 1. Atribut Dan Penjualan Produk Sepatu Pada PT. Zoelindo Elite Anugerah**

No	Atribut	Nilai
1	Model	Casual, Flatshoes, 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, Javelin, 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

##### 4.b. Seleksi Data

Pemilihan atribut atau variabel tersebut dengan pertimbangan bahwa jumlah nilai variabelnya tidak banyak sehingga diharapkan minat pembeli masuk kedalam satu klasifikasi nilai variabel tersebut cukup banyak. Dari data yang ada, variabel diambil sebagai atribut atau variabel dalam pembentukan pohon keputusan adalah: Model, Warna, Kelamin, Size, Bahan, Merk, dan Kategori.

##### 4.c. Transformasi Data

Transformasi data dilakukan dengan mengubah beberapa nilai atribut yang awalnya bernilai angka-angka menjadi nilai atribut yang sesuai dengan data pada tabel 2, agar dapat dilakukan proses perhitungan algoritma klasifikasi C4.5.

##### 4.d. Perhitungan Entropy dan Information Gain

Dengan merubah rumus (1), kita dapat menghitung nilai *Entropy*. Menghitung *Entropy* Total dilakukan dengan cara menghitung jumlah keputusan "Memenuhi" dan "Tidak Memenuhi".

$$- Entropy (Total) = (-31/35 \cdot \log_2 (31/35)) + (-4/35 \cdot \log_2 (4/35)) = 0,1140$$

*Entropy* Total adalah menghitung nilai total keputusan "Memenuhi" (31) dan "Tidak Memenuhi" (4), sedangkan 35 adalah jumlah keseluruhan kasus. Kemudian menghitung masing-masing *Entropy* dari semua nilai variabel yang ada. Dan dengan menggunakan rumus agar kita dapat menghitung nilai *Information Gain* tiap atribut.

**Tabel 2. Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah**

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

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

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

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

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

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

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

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

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

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

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

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

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

$$E_{putih} [6,13] = (-6/19 \log_2 6/19) + (-13/19 \log_2 13/19)$$

$$13/19) = 0,8997$$

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

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

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

Nilai Entropy nya dapat dihitung sebagai berikut:

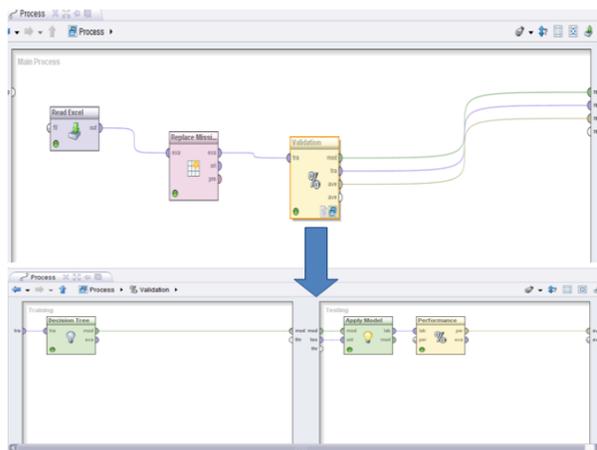
$$Entropy\ split = \sum_{i=1}^p \frac{n_i}{n} IE(i)$$

E split warna

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

$$Gain\ warna = 0,9469 - 0,8091 = 0.1378$$

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



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

Langkah awal pengolahnya 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.

Dari hasil uji coba menggunakan algoritma C4.5 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.

**Evaluasi dan Validasi Hasil**

Pengujian model dengan C4.5 dengan menentukan nilai *accuracy* dan AUC, untuk menghasilkan Memprediksi Minat Pelanggan Membeli Produk Sepatu. Dalam menentukan tingkat keakurasian dalam model C4.5. Berikut adalah desain model Metode pengujiannya menggunakan *cross validation*.

**Hasil Pengujian Model Algoritma C4.5**

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

a. *Confusion Matrix*

Tabel 4.1 adalah perhitungan data *training* pada Table 4.4 dengan jumlah data 490 record. Berikut Tabel yang didapat:

**Tabel 5 Model Confusion Matrix untuk Algoritma C4.5**

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	66,04%
class recall	82,32%	59,22%	

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

	%
Accuracy	73.88
Sensitivity	77.81
Specificity	65.84
PPV	82.32
Npv	59.22

**Tabel 6.** Nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv*

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$= \frac{256 + 106}{256 + 106 + 55 + 73} = \frac{362}{490} = 0,7387$$

$$= 0,7387 \times 100\% = 73,88 \%$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$= \frac{256}{256 + 73} = \frac{256}{329} = 0,7781$$

$$= 0,7781 \times 100\% = 77,81\%$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$= \frac{106}{106 + 55} = \frac{106}{161} = 0,6584$$

$$= 0,6584 \times 100\% = 65,84 \%$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$= \frac{256}{256 + 55} = \frac{256}{311} = 0,8232$$

$$= 0,8232 \times 100\% = 82,32\%$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

$$= \frac{106}{106 + 73} = \frac{106}{179} = 0,5922$$

$$= 0,5922 \times 100\% = 59,22\%$$

#### b. Evaluasi ROC Curve

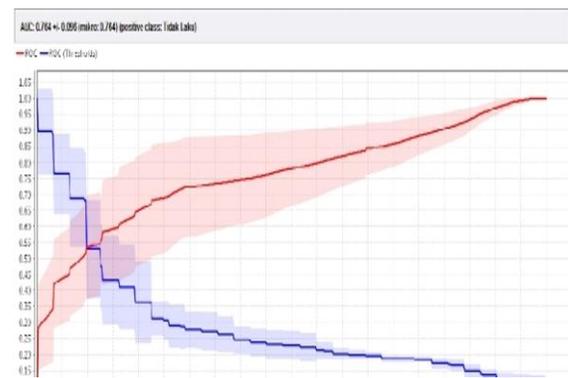
Dari Gambar 5 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.780. Tingkat akurasi dapat di diagnosa sebagai berikut:

Akurasi 0.90 - 1.00 = *Excellent classification*

Akurasi 0.80 - 0.90 = *Good classification*

	Laki	Tidak Laki
Warna	170	1
Bahan Upper	1	1
Merk	120	1
Gelang	120	1

Akurasi 0.70 - 0.80 = *Fair classification*



Dari gambar 5 terdapat gambar grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.764 dengan diagnosa hasilnya *Fair classification*

Dari hasil uji coba menggunakan algoritma C4.5 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.

#### 4.g. Validasi dan Pengujian

Pengujian dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah 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 membandingkan hasil akurasi dan AUC, sehingga diperoleh model dari metode klasifikasi yang mana yang memperoleh nilai akurasi dan AUC tertinggi:

```

PerformanceVector
PerformanceVector:
accuracy: 90.83% +/- 14.17% (mikro: 91.43%)
ConfusionMatrix:
True: Memenuhi          Tidak_Memenuhi
Memenuhi: 31             3
Tidak_Memenuhi: 0       1
classification_error: 9.17% +/- 14.17% (mikro: 8.57%)
ConfusionMatrix:
True: Memenuhi          Tidak_Memenuhi
Memenuhi: 31             3
Tidak_Memenuhi: 0       1
kappa: 0.371
ConfusionMatrix:
True: Memenuhi          Tidak_Memenuhi
Memenuhi: 31             3
Tidak_Memenuhi: 0       1
    
```

**Gambar 3. Hasil validasi dilakukan pengujian menggunakan k-fold cross validation**

Berdasarkan pengujian dengan menggunakan metode *k-fold cross validation* menghasilkan nilai akurasi yang hampir sempurna sebesar 90,83 %, presisi sebesar 91,18% dan recall sebesar 62,50 %. Hal ini menunjukkan bahwa, dari proses klasifikasi yang dilakukan akan dapat diterapkan pada rekomendasi Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah.

**4.h. Analisis Hasil Pengujian**

Analisis hasil pengujian dilakukan dengan melakukan perhitungan secara manual dengan *Confusion Matrix*. Perhitungan menggunakan model *confusion matrix*. Berikut ini merupakan hasil dari perhitungan *confusion matrix* pada algoritma C4.5:

**Tabel 4. Hasil Pengujian**

	<i>Accuracy</i>	<i>AUC</i>
<b>C4.5</b>	73.88%	0.764

Dari Perhitungan diatas, dapat disimpulkan bahwa hasil dari perhitungan *accuracy*, *Sensitivity*, *Specificity*, *PPV*, dan *Npv* tersebut sama dengan hasil perhitungan yang ditampilkan pada Tabel 6. Berdasarkan pengujian dan analisa hasil pengujian yang dilakukan, dengan tingkat akurasi 73.88%, *Sensitivity* 77.81 %, *Specificity* 65.84 % , *PPV* 82.32 % dan *Npv* 59.22 % menunjukkan nilai yang hampir seratus persen akurat, dan selebihnya yang masih dalam kategori baik menyimpulkan bahwa peneliti berhasil dalam mengimplementasikan algoritma klasifikasi C4.5 dengan baik dan akan membantu PT. Zoelindo Elite Anugerah dalam Memprediksi Minat Pelanggan Membeli Produk Sepatu.

**5. Kesimpulan**

Dari penelitian yang dilakukan adalah sebagai berikut:

Dari hasil analisis optimasi model algoritma C4.5 memberikan nilai akurasi yang lebih tinggi 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*., Dari evaluasi *ROC curve* tersebut terlihat bahwa model algoritma C4.5 cukup baik Sehingga dapat disimpulkan bahwa algoritma C4.5 cukup akurat dalam memprediksi minat beli prodak sepatu.

**6. Pustaka**

- [1] Arifin. (2015). Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur. Jurnal Informatika. Vol.II. No.2
- [2] Defiyanti, Sofi. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. Jurnal CITEE, 257.
- [3] Harryanto, F.F, & Hansun, Seng. (2017). Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. Jurnal Jatisi, Vol 3, 95-103.
- [4] Hasibuan, Malayu S.P. 2006, Manajemen Sumber Daya Manusia, edisi Revisi. Jakarta. Bumi Aksara.
- [5] Iriadi, Nandang & Nuraeni. (2016). Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta: Jurnal Teknik Komputer Amik BSI. Vol.II. No.1.
- [6] Lumbantoruan & Kennedy. (2015). Analisis Data Mining Dan Warehousing: Jurnal Ilmiah Buletin Ekonomi. Vol.19. No.1.
- [7] Sulistiyani, T., & Ambar, R. (2003). Manajemen Sumber Daya Manusia, Yogyakarta: Graha Ilmu.