Perbandingan Algoritma Klasifikasi *Naive Bayes*, *Nearest Neighbour*, dan *Decision Tree* pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian

Dewi Sartika*[1], Dana Indra Sensuse*[2]
Universitas Indo Global Mandiri
Fakultas Ilmu Komputer
dewi.sartika31@alumni.ui.ac.id*[1], dana@cs.ui.ac.id*[2]

Abstract

Data mining is a process of analysis of the large data set in the database so that the information obtained will be used for the next stage. One technique commonly used data mining is the technique of classification. Classification is an engineering modeling of the data that has not been classified, to be used to classify new data. Classification included into any type of supervised learning, meaning that it takes the training data to build a model of classification. There are five categories of classification that is statistically based, distance-based, based on the decision tree, neural network-based and rule-based. Each category has many options classification algorithms, some algorithms are frequently used algorithms Naive Bayes, nearest neighbor and decision tree. In this study will be a comparison of the three algorithms on case studies of electoral decision making clothing patterns. The comparison showed that the decision tree method has the highest level of accuracy than Naive Bayes algorithm and nearest neighbor, reaching 75.6%. Decision tree algorithm used is J48 with pruned algorithm that produces models of decision tree with leaves as many as 166 and 255 magnitude decision tree.

Keywords: *Supervised Learning, Naive Bayes, Nearest Neighbor, Decision Tree, J48.*

Abstrak

Data mining adalah suatu proses analisis terhadap sekumpulan data yang ada di dalam basis data sehingga diperoleh informasi yang akan digunakan untuk tahap selanjutnya. Salah satu teknik data mining yang umum digunakan yaitu teknik klasifikasi. Klasifikasi adalah suatu teknik pembentukan model dari data yang belum terklasifikasi, untuk digunakan mengklasifikasi data baru. Klasifikasi termasuk ke dalam tipe supervised learning, artinya dibutuhkan data pelatihan untuk membangun suatu model klasifikasinya. terdapat 5 kategori klasifikasi yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan syaraf, dan berbasis aturan. Tiap kategori klasifikasi memiliki banyak pilihan algoritma, beberapa algoritma yang sering digunakan adalah algoritma naive bayes, nearest neighbour, dan decision tree. Pada penelitian ini akan dilakukan perbandingan dari ketiga algoritma tersebut pada studi kasus pengambilan keputusan pemilihan pola pakaian. Hasil perbandingan menunjukkan bahwa metode decision tree memiliki tingkat akurasi tertinggi dibandingkan algoritma naive bayes dan nearest neighbour yaitu mencapai 75.6%. Algoritma decision tree yang digunakan ialah algoritma J48 dengan pruned yang menghasilkan model decision tree dengan daun sebanyak 166 dan pohon keputusan yang besarnya 255.

Kata kunci**:** *Supervised Learning, Naive Bayes,* N*earest Neighbour,* D*ecision Tree,* J48.

1. PENDAHULUAN

Data mining adalah proses yang memanfaatkan suatu metode untuk memperoleh pola dari suatu data [1], sedangkan menurut Mirkin *Data mining* didefinisikan sebagai suatu proses untuk mencari pola dari sekumpulan data yang terdapat di dalam *database* untuk kemudian dianalisis sehingga menghasilkan suatu informasi tertentu untuk dimanfaatkan pada proses selanjutnya [2]. Salah satu pendekatan yang dapat digunakan untuk menganalisis sekumpulan data adalah klasifikasi. Klasifikasi merupakan salah satu teknik data mining yang digunakan untuk membangun suatu model dari sampel data yang belum terklasifikasi untuk digunakan mengklasifikasi sampel data baru ke dalam kelas-kelas yang sejenis [3]. Klasifikasi termasuk ke dalam *supervised learning* karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu, kemudian pola dari hasil analisis tersebut digunakan untuk pengklasifikasian data uji. Proses klasifikasi data terdiri dari pembelajaran dan klasifikasi. Pada pembelajaran data *training* dianalisis menggunakan algoritma klasifikasi, selanjutnya pada klasifikasi digunakan data *testing* untuk memastikan tingkat akurasi dari *rule* klasifikasi yang digunakan. Teknik klasifikasi dibagi menjadi lima kategori berdasarkan perbedaan konsep matematika, yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan syaraf, dan berbasis *rule* [4]. Ada banyak algoritma dari masing-masing kategori tersebut, namun yang populer dan sering digunakan diantaranya yaitu *naive bayes*, *nearest neighbour* dan *decision tree*. Pada penelitian ini peneliti akan mencoba membandingkan algoritma klasifikasi *nearest neighbor*, *naive bayes* dan *decision tree* menggunakan tools WEKA. Ketiga metode tersebut akan dibandingkan berdasarkan tingkat akurasinya. Studi kasus yang digunakan pada penelitian ini adalah pengambilan keputusan pemilihan pola pakaian, dimana datanya berupa kasus-kasus terdahulu yang telah memiliki label kelas solusi pola. Kasus merupakan sekumpulan informasi atribut berupa kriteria-kriteria yang digunakan dalam pengambilan keputusan pemilihan pola pakaian. Kriteria-kriteria tersebut diperoleh dari wawancara *expert*. Perbedaan studi kasus pada penelitian ini dengan studi kasus pada penelitian-penelitian terdahulu adalah banyaknya label kelas yang digunakan dalam pengklasifikasian data.

2. METODE PENELITIAN

2.1. Data Mining

Saat ini kebutuhan terhadap analisis suatu data sangat dibutuhkan. Perkembangan jumlah data yang semakin pesat mendorong untuk memanfaatkannya dalam penggalian informasi maupun pengetahuan. Seperti sekumpulan data transaksi pada supermarket dapat dianalisis sehingga diperoleh informasi terkait barang-barang apa saja yang sering dibeli oleh konsumen, dengan mengetahui informasi tersebut dapat membantu dalam penentuan jumlah stok barang untuk selanjutnya, dengan begitu diharapkan dapat meningkatkan keuntungan yang diperoleh dan penumpukan stok barang karena supermarket mampu menentukan stok barang dengan tepat sesuai dengan minat dan kebutuhan konsumen. *Data mining* diartikan sebagai penentuan pola dari hasil analisis sekumpulan data [2]. Data mining juga dapat diartikan sebagai suatu proses logikal yang digunakan untuk mencari dari sejumlah data untuk mendapatkan data yang berguna [3].

Sebelum menerapkan algoritma *data mining*, harus dipahami terlebih dahulu bahwa algoritma *data mining* terbagi dalam dua kategori yaitu deskriptif dan prediktif. Deskriptif adalah mengukur kesamaan antar objek serta menemukan pola dan hubungan yang belum diketahui di dalam sekumpulan data, sedangkan prediktif adalah penyimpulan *rule* prediksi dari data pelatihan, yang selanjutnya *rule* tersebut digunakan pada data yang belum terprediksi [9]. Beberapa fungsi *data mining* yang termasuk deskriptif adalah *clustering*, *summarization*, dan

sequence discover, sedangkan fungsi *data mining* yang termasuk prediktif adalah klasifikasi, regresi, *time series analysis*, dan *prediction.*

2.2. Klasifikasi

Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas [1]. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas dalam statistik [9]. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data. Contoh sederhana dari teknik *data mining* klasifikasi adalah pengklasifikasian hewan berdasarkan atribut jumlah kaki, habitat dan organ pernafasannya akan diklasifikasikan ke dalam dua label kelas yaitu unggas dan ikan. Label kelas unggas adalah data yang memiliki jumlah kaki dua, habitatnya di darat, dan organ pernafasannya menggunakan paru-paru, sedangkan label kelas ikan adalah data yang memiliki jumlah kaki nol (tidak memiliki kaki), habitat di air, dan organ pernafasannya menggunakan insang. Banyak algoritma yang dapat digunakan dalam pengklasifikasian data, namun dalam penelitian ini hanya akan membandingkan tiga algoritma saja, yakni *naive bayes*, *nearest neighbour*, dan *decision tree.*

2.2.1. Naive Bayes

Teorema bayes adalah perhitungan statistik dengan menghitung probabilitas kemiripan kasus lama yang ada dibasis kasus dengan kasus baru. *Teorema bayes* memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada *database* yang besar [1]. *Naive bayes* termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut. *Naive bayes* merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma *naive bayes* juga mampu menangani set data yang memiliki banyak atribut [9]. Persamaan dari *naive bayes* sebagai berikut:

$$P(C_i \mid X) = \frac{P(X \mid C_i)P(C_i)}{P(X)} \quad (1)$$

Keterangan :

X	: Kriteria suatu kasus berdasarkan masukan
C_i	: Kelas solusi pola ke-i, dimana i adalah jumlah label kelas
$P(C_i\|X)$	: Probabilitas kemunculan label kelas C_i dengan kriteria masukan X
$P(X\|C_i)$	: Probabilitas kriteria masukan X dengan label kelas C_i
$P(C_i)$	: Probabilitas label kelas C_i

2.2.2. Nearest Neighbour

Nearest Neighbour adalah algoritma pengklasifikasian yang didasarkan pada analogi, yaitu membandingkan data uji dengan data pelatihan yang berada dekat dengan dan memiliki kemiripan dengan data uji tersebut [10]. Kemiripan data uji dengan data pelatihan didasarkan pada jaraknya. Banyak persamaan yang dapat digunakan untuk menghitung jarak antara data uji dan data pelatihan. Tiga diantaranya yang paling sering digunakan adalah:

1. Atribut yang bertipe numerik

Terdapat dua pendekatan perhitungan jarak/kemiripan yang umum digunakan untuk atribut yang bertipe numerik, yaitu *euclidean distance* [10] dengan persamaan berikut:

$$Dist(x1, x2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})} \quad (2)$$

Keterangan:
n : jumlah data
x_1 : data uji
x_2 : data pembelajaran

Persamaan yang kedua yaitu *Manhattan distance* [11] sebagai berikut :

$$Dist(p_i(an), p_i(nc)) = \frac{p_i(an) - p_i(nc)}{\max\_ dist\_ i} \quad (3)$$

Keterangan:
p_i : atribut ke-i
an : data pembelajaran
nc : data uji

2. Atribut yang bertipe simbolik

Persamaan yang digunakan untuk atribut yang menggunakan istilah eksplisit yaitu ada atau tidak ada, memiliki atau tidak memiliki, ya atau tidak dan sebagainya maka perhitungan kemiripan atau jarak dapat dihitung dengan fungsi sebagai berikut [12]:

$$Sim(Ki(a), Ki(b)) = \begin{cases} 0 & Ki(a) \neq Ki(b) \\ 1 & Ki(a) = Ki(b) \end{cases} \quad (4)$$

Keterangan :
Ki(a) : kriteria ke-i dari kasus a
Ki(b) : kriteria ke-i dari kasus b
Sim(Ki(a),Ki(b) : nilai kemiripan kriteria ke-i antara kasus a dengan kasus b

Perhitungan selanjutnya adalah persamaan untuk mencari kemiripan dengan *nearest neighbour* yaitu [13]:

$$Similarity(T, S) = \frac{\sum_{i=1}^{n} Sim(K_i(T), K_i(S)) x w_i}{\sum_{i=i}^{n} w} \quad (5)$$

Keterangan:
T : data uji
S : data pembelajaran
n : jumlah kriteria
w : bobot kriteria
$Sim(K_i(T), K_i(S))$: Nilai kemiripan/jarak kriteria kasus target dan target sumber

2.2.3. Decision Tree

Algotirma *decision tree* merupakan algoritma yang umum digunakan untuk pengambilan keputusan. *Decision tree* akan mencari solusi permasalahan dengan menjadikan kriteria sebagai *node* yang saling berhubungan membentuk seperti struktur pohon [14]. *Decision tree* adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon [8]. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (tidak ada cabang lagi). Konsep data dalam

decision tree adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan *record.* Atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon.

Proses dalam *decision tree* adalah sebagai berikut [7]:

1. Mengubah bentuk data (tabel) menjadi model pohon

Hal yang dilakukan pada tahapan ini adalah menentukan atribut yang terpilih mulai dari akar, cabang hingga menuju keputusan. Banyak pendekatan yang dapat digunakan untuk menentukan atribut terpilih, pada penelitian ini akan menggunakan perhitungan *gainratio* dari setiap kriteria dengan data sampel. Untuk menghitung nilai *gainratio* dapat dilakukan dengan persamaan sebagai berikut:

$$Gainratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (6)$$

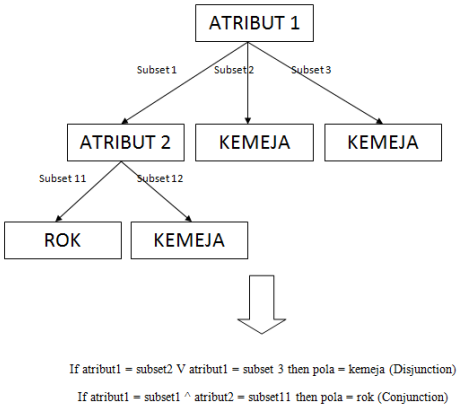
Dimana nilai *information gain* bermakna seberapa banyak informasi yang diperoleh dengan mengetahui nilai suatu atribut sedangkan nilai *split information* digunakan untuk suatu atribut yang memiliki banyak *instance* (lebih dari dua dan beragam)

2. Mengubah model pohon menjadi *rule*

Formula untuk membangkitkan *rule* didefinisikan sebagai berikut:

IF premis THEN konklusi (7)

Simpul akar dan cabang akan menjadi premis dari aturan, sedangkan simpul daun akan menjadi bagian dari konklusinya (solusi). Tiap premis yang terdapat dalam satu atribut akan dihubungkan dengan hubungan disjungsi, sedangkan premis yang memiliki lanjutan premis pada cabang selanjutnya akan dihubungkan dengan konjungsi.



Gambar 1. Proses Model Pohon Menjadi *Rule*

3. Menyederhanakan *rule* (*Pruning*)

Pada proses penyederhanaan *rule*, tahapan-tahapan dilakukan sebagai berikut:

1. Membuat tabel distibusi terpadu dengan menyatakan semua nilai kejadian pada setiap *rule*.
2. Menghitung tingkat *independensi* antara kriteria pada suatu *rule*, yaitu antara atribut dengan target atribut (perhitungan tingkat *independensi* menggunakan *test of independency Chi-Square*).
3. Mengeliminasi kriteria yang dianggap tidak perlu, yaitu yang memiliki tingkat *independensi* tinggi.

Misalkan yang ingin dilihat adalah pengaruh jenis pakaian terhadap penentuan solusi pola pakaian yang dapat dibuat, tentukan terlebih dahulu tingkat signifikansinya ($\lambda = 0.05$), sehingga dapat dihitung *degree of freedom* dengan persamaan berikut :

$$\{0.05;(r-1)*(c-1)\} \quad (8)$$

Keterangan:

r : jumlah baris
c : jumlah kolom
Setelah diperoleh nilainya maka dapat dilihat pada tabel untuk memperoleh nilai X^2_{tabel} untuk dibandingkan dengan X^2_{hitung}. X^2_{hitung} diperoleh melalui persamaan berikut :

$$X^2_{hitung} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (9)$$

Keterangan:
n_{ij} : nilai *record* baris ke i kolom ke j dari tabel distribusi terpadu.
Sedangkan nilai e_{ij} diperoleh melalui persamaan berikut:

$$e_{ij} = \frac{n_i \bullet n_j}{n} \quad (10)$$

Keterangan:
n_i : marjinal dari baris ke i
$n_{j:}$ marjinal dari kolom ke j
n : jumlah *record* data

Jika nilai $X^2_{hitung} \leq X^2_{tabel}$ artinya atribut tersebut tidak mempengaruhi atribut target, sehingga *rule* dari atribut tersebut dapat dihilangkan. Namun sebaliknya jika nilai $X^2_{hitung} > X^2_{tabel}$ berarti atribut tersebut mempengaruhi atribut target, sehingga *rule* dari atribut tersebut tidak dapat dihilangkan.

2.3. WEKA

Waikato Environment for Knowledge Analysis (WEKA) merupakan perangkat lunak pembelajaran mesin yang populer yang ditulis dalam bahasa pemrograman java. WEKA dikembangkan di Universitas Waikato, Selandia Baru. WEKA berisikan kumpulan algoritma beserta visualisasinya untuk analisis data dan pemodelan prediktif. Algoritma-algoritma pembelajaran mesin pada WEKA dapat dimanfaatkan untuk pemecahan masalah dibidang *data mining*. WEKA versi asli awalnya dirancang untuk menganalisis data dari domain pertanian, tetapi WEKA versi lengkap berbasis java (versi 3), yang mulai dibangun pada tahun 1997, yang sekarang dapat digunakan untuk menganalisis data dari berbagai domain, khususnya untuk pendidikan dan penelitian. WEKA memiliki implementasi semua teknik pembelajaran untuk klasifikasi dan regresi , yaitu *decision trees*, *rules set*, pengklasifikasian *teorema bayes*, *Support Vector Machines* (SVM), logistik dan linier, *multi layers perceptrons* dan metode *nearest neighbour* [15]

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Algoritma

Perangkat Lunak WEKA menyediakan 4 macam mode pengujian [19], yaitu:

1. *Use Training Set*: mengevaluasi seberapa baik algoritma mampu memprediksi kelas dari *instance* setelah dilakukan pelatihan. Data pelatihan akan digunakan untuk data uji.
2. *Supplied Test*: mengevaluasi seberapa baik algoritma mampu memprediksi kelas dari set *instance* yang diambil dari suatu file. Data pelatihan dan data uji berbeda file.
3. *Cross-validation*: mengevaluasi algoritma melalui *cross-validation*, menggunakan nilai *folds* yang dimasukkan.
4. *Percentage split*: mengevaluasi seberapa baik algoritma mampu memprediksi persentase tertentu dari data. Data set akan dibagi menjadi 2, data pelatihan dan data uji.

Mode pengujian yang dilakukan pada penelitian ini adalah mode 1, 3 dan 4. Evalusai pengujian yang dilakukan terhadap algoritma klasifikasi *naive bayes*, *nearest neighbour*, dan *decision tree* dalam penelitian ini yaitu seberapa besar tingkat akurasi yang dihasilkan algoritma

dalam memprediksi solusi pola pakaian yang tepat. Pola pakaian yang tepat merupakan solusi pola pakaian yang sama dengan keputusan solusi pola pakaian dari *expert*. Persentase tingkat akurasi akan dihitung dengan persamaan sebagai berikut:

$$PA = \frac{PB}{TB} x100\% \quad (11)$$

Keterangan:
PA : Persentasi Akurasi
PB : Total prediksi benar
TB : Total *instance* keseluruhan

3.2. Perbandingan Decision Tree Pruned dan Unpruned

Pemangkasan (*pruned*) dalam algoritma klasifikasi *decision tree* merupakan tahapan fundamental dalam mengoptimalkan efisiensi komputasi beserta akurasi pengklasifikasian [20]. Model pohon sebelum dan sesudah pemangkasan akan diuji dengan mode pengujian *percentage split* dengan persentasi data pelatihan sebanyak 90, 80, 70 dan 60. Hasil perbandingan tingkat akurasi yang dihasilkan oleh pohon yang terpangkas dan tidak dapat dilihat pada tabel 1:

Tabel 1. Hasil Perbandingan *Pruned* dan *Unpruned Tree*

Data Pelatihan	***Unpruned***			***Pruned***		
	Prediksi Benar	**Prediksi Salah**	**Akurasi (%)**	**Prediksi Benar**	**Prediksi Salah**	**Akurasi (%)**
90	31	10	75.6	31	10	75.6
80	57	25	69.5	58	24	70.7
70	70	53	56.9	75	48	60.9
60	92	72	56	94	70	57.3

Berdasarkan hasil yang diperoleh pada Tabel 2 bahwa pohon keputusan yang dipanggas menghasilkan tingkat akurasi yang lebih tinggi dibandingkan pohon keputusan yang tidak dipanggas.

3.3. Eksperimen Pengujian

3.3.1. Mode Pengujian Use Training Test

Pada mode pengujian ini semua data set sejumlah 408 data akan dijadikan data pelatihan dan data pengujian. Hasil dari mode pengujian ini akan mencapai nilai tingkat akurasi yang besar hingga mencapai 100%, dikarenakan data pelatihan dan data uji yang digunakan adalah data yang sama. Hasil dari mode pengujian ini dapat dilihat pada tabel berikut:

Tabel 2. Hasil Pengujian Mode *Use Training Test*

No	**Algoritma Klasifikasi**	**Prediksi Benar**	**Prediksi Salah**	**Akurasi (%)**
1	*Naive Bayes*	367	41	89.9
2	IB1	408	0	100
3	J48	368	40	90.1

Pada Tabel 2 memperlihatkan bahwa pada mode pengujian ini, algoritma IB1 unnggul dengan tingkat akurasi mencapai 100%, dari hasil analisis hal ini disebabkan algoritma IB1 hanya melakukan perhitungan kemiripan berdasarkan ekslipit suatu atribut kriteria, sehingga nilai kemiripan antara data uji dan data training akan bernilai 1 karena menggunakan data yang sama persis.

3.3.2. Mode Pengujian Percentage Split

Pada mode penelitian ini data set sebanyak 408 data akan dibagi menjadi data pelatihan dan data uji sesuai persentase yang ditentukan. Persentase yang digunakan dalam mode pengujian ini adalah dengan data pelatihan sebesar 90%, 80%, 70% dan 60% dari data set. Hasil dari pengujian mode ini pada algoritma *naive bayes, nearest neighbour* dan *decision tree* dapat dilihat pada tabel berikut:

Tabel 3. Hasil Pengujian Mode *Percentage Split*

Data Training (%)	Naive Bayes			IB1			J48		
	Prediksi Benar	Prediksi Salah	Akurasi (%)	Prediksi Benar	Prediksi Salah	Akurasi (%)	Prediksi Benar	Prediksi Salah	Akurasi (%)
90	17	24	41.5	8	33	19.5	31	10	75.6
80	32	50	39	16	66	19.5	58	24	70.7
70	39	84	31.7	25	98	20.3	75	48	60.9
60	44	120	26.8	37	127	22.5	94	70	57.3

3.3.3. Mode Pengujian Supplied Test

Pada pengujian mode ini akan digunakan data uji diluar dari data set. Data uji tersebut belum memiliki solusi pola pakaian dari *expert.* Guna dilakukannya mode pengujian ini adalah melihat apakah algoritma klasifikasi *naive bayes, nearest neighbour* dan *decision tree* mampu memberikan prediksi pola pakaian. Sebaran sub kriteria pilihan untuk data uji adalah semua sub kriteria pilihan dari ketujuh kriteria akan muncul, sehingga total data uji sebanyak 13 data dengan detail sebagai berikut:

Tabel 4. Data Uji Pengujian *Supplied Test*

No	Jenis Pakaian	Kategori	Jenis Bahan	Ukuran Bahan (meter)	Ukuran Badan	Jenis Kelamin	Usia
1	Formal	Atasan	Twistone	2.5	Extra Large	Laki-laki	Dewasa
2	Non Formal	Celana	Katun Linen	2.5	Large	Laki-laki	Dewasa
3	Non Formal	Jumpsuit	Valvet	3	Large	Wanita	Dewasa
4	Non Formal	Luaran	Saten	1	Small	Wanita	Dewasa
5	Formal	Kemeja	Drill	1.5	Small	Wanita	Dewasa
6	Formal	Atasan	Wedges	2	Small	Wanita	Dewasa
7	Non Formal	Baju Tidur	Katun	2	Medium	Laki-laki	Dewasa
8	Non Formal	Baju Muslim	Kaos	3	Medium	Wanita	Dewasa
9	Formal	Rok	Flanel	3.5	Extra Large	Wanita	Dewasa
10	Non Formal	Rok	Denim	1	Large	Wanita	Anak-anak
11	Non Formal	Dress	Katun Rayon	1	Medium	Wanita	Anak-anak
12	Formal	Kemeja	Batik	1	Medium	Laki-laki	Anak-anak
13	Non Formal	Luaran	Spandex	1	Small	Wanita	Anak-anak

Hasil dari pengujian mode *supplied test* dengan data uji seperti pada Tabel 3 dan data pelatihan menggunakan seluruh data set yaitu sebanyak 408 data. Hasil analisis yang diperoleh dari hasil pengujian mode *supplied test* yang ditampilkan pada Tabel 4 bahwa:

1. *Naive Bayes*: memberikan prediksi pola pakaian yang kurang tepat pada data 3,5,8,9 karena tidak sesuai kategori, sedangkan pada data 7 tidak sesuai dengan jenis kelamin, dan pada data 12 tidak sesuai dengan usia.
2. IB1: memberikan prediksi pola pakaian yang kurang tepat pada data 3,5,6,8 karena tidak sesuai kategori, sedangkan pada data 1 dan 7 tidak sesuai dengan jenis kelamin, dan pada data 12 tidak sesuai dengan usia.
3. J48: prediksi pola pakaian yang kurang tepat pada data 1,2,7 karena tidak sesuai dengan jenis kelamin, sedangkan pada data 10,11,12 tidak sesuai dengan usia.

Tabel 5. Hasil Pengujian Mode *Supplied Test*

No	**Algoritma Klasifikasi**		
	Naive Bayes	**IB1**	**J48**
1	Baju Koko	Atasan Wanita Lengan Panjang	Atasan Wanita Lengan Panjang
2	Kulot Panjang Laki-laki	Kulot Panjang Laki-laki	Kulot Panjang Wanita
3	Rok Duyung Panjang	Rok Duyung Panjang	Jumpsuit Can see Panjang
4	Rompi Berkerah	Bolero Lengan A	Rompi Kutung
5	Rompi Kutung	Celana Panjang Wanita	Kemeja Lengan Panjang
6	Atasan Wanita Formal	Dress Pendek	Atasan Peplum Lengan Pendek
7	Daster Can See	Daster Can See	Daster Can See
8	Rok Tulip Panjang	Atasan Lengan Batwing	Gamis Lipat
9	Gamis Lipat	Rok A Panjang	Rok Lipat Pendek
10	Rok Lipat Anak	Rok Lipat Anak	Rok Lipat Pendek
11	Dress Can see Anak	Dress Can see Anak	Dress Pendek
12	Kemeja Laki-laki Lengan Pendek	Kemeja Laki-laki Lengan Pendek	Kemeja Lengan Panjang
13	Blezer Anak	Blezer Anak	Rompi Kutung

4. KESIMPULAN

Kesimpulan yang diperoleh pada penelitian ini adalah dari hasil perbandingan algoritma klasifikasi *nearest neighbour*, *naive bayes* dan *decision tree* yang digunakan pada studi kasus pengambilan keputusan pemilihan pola pakaian menyatakan bahwa algoritma klasifikasi *decision tree* merupakan algoritma klasifikasi yang memiliki tingkat akurasi paling tinggi dibandingkan algoritma klasifikasi *naive bayes* dan *nearest neighbour* yaitu mencapai 75.6% pada pengujian yang dilakukan dengan menggunakan mode pengujian *percentage split*. Algoritma *decision tree* yang digunakan ialah algoritma J48 dengan *pruned* yang model pohon keputusannya memiliki 166 daun dan ukuran pohon 225.

5. SARAN

Berdasarkan hasil temuan pada implikasi penelitian, maka saran untuk penelitian selanjutnya adalah :

1. Penggunaan *multiple expert* dapat diterapkan untuk memperolah kriteria-kriteria yang lebih bervariatif dan jumlah kasus lebih banyak.
2. Menspesifikasikan keriteria-keriteria yang dibuat umum, seperti ukuran bahan dan ukuran badan.

3. Melanjutkan penelitian ini dengan mengimplementasikan metode yang terbaik sistem pendukung keputusan pemilihan pola pakaian.

DAFTAR PUSTAKA

[1] J. Iawe. Han, M. Kamber, and J. Pei, 2012, *Data Mining Concept and Techniques*.

[2] B. Mirkin, 2011, "*Data Analysis, Mathematical Statistics, Machine Learning, Data Mining: Similarities and Differences*," *2011 Int. Conf. Adv. Comput. Sci. Inf. Syst.*, Vol. 2, pp. 1–8.

[3] M. Ramageri, 2010, "*Data Mining Techniques and Applications*," *Indian J. Comput. Sci. Eng.*, Vol. 1, No. 4, pp. 301–305.

[4] M. Karim and R. M.Rahman, 2013, "*Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing*," *J. Softw. Eng. Appl.*, Vol. 6, pp. 196–206.

[5] F. Octaviani S, J. Purwadi, and R. Delima, "*Implementasi Case Based Reasoning untuk Sistem Diagnosis Penyakit Anjing*."

[6] I. A. Mubarok, N. A. Wesiani, and A. Rusdiansyah, "*Pengembangan Prototype Knowledge Management System berbasis Case Based Reasoning bagi Peningkatan Aksesibilitas UMKM Dalam Permodalan Usaha*," pp. 1–6.

[7] V. Mandasari and B. A. Tama, 2011, "*Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining: Studi Kasus XYZ*," *J. Generic*, Vol. 6, No. 1, pp. 25–28.

[8] N. Jayanti, S. Puspitodjati, and T. Elida, 2008, "*Teknik Klasifikasi Pohon Keputusan untuk Memprediksi Kebangkrutan Bank Berdasarkan Ratio Keuangan Bank*," pp. 101–107.

[9] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J.-F. Chang, and L. Hua, 2012, "*Data Mining in Healthcare and Biomedicine: A Survey of The Literature*," pp. 2431–2448,

[10] R. Entezari-Maleki, A. Rezaei, and B. Minaei-Bidgoli, "*Comparison of Classification Methods Based on The Type of Attributes and Sample Size*."

[11] S. Guessoum, M. T. Laskri, and M. T. Khadir, 2012, "*Combining Case and Rule Based Reasoning for The Diagnosis and Therapy of Chronic Obstructive Pulmonary Disease*," *Int. J. Hybrid Inf. Technol.*, vol. 5, no. 3, pp. 145–160.

[12] D. Gu, C. Liang, X. Li, S. Yang, and Z. Pei, 2010, "*Intelligent Technique for Knowledge Reuse of Dental Medical Records Based on Case-Based Reasoning*," pp. 213–222.

[13] S. H. Kang and S. K. Lau, 2002, "*Intelligent Knowledge Acquisition with Case-Based Reasoning Techniques 2. CASE-BASED REASONING SYSTEMS TECHNIQUES Technical Report 2002 May 2002*".

[14] S. H. Babic, P. Kokol, V. Podgorelec, M. Zorman, M. Sprogar, and M. M. Stiglic, 2000, "*The Art of Building Decision Trees*," *J. Med. Syst.*, Vol. 24, No. 1, pp. 43–52.

[15] E. Frank, M. Hall, L. Trigg, G. Holmes, and I. H. Witten, 2003, "*Data Mining in Bio Informatics Using Weka*," pp. 1–2.

[16] F. Afiff, "*Kewirausahaan dan Ekonomi Kreatif*," 2012, [Online]. Available: http://id.wikipedia.org/wiki/Ekonomi_kreatif#cite_note-ekraf8-8. [Accessed: 27-May-2015].

[17] E. Siregar and F. Hutapea, "*Perbedaan Hasil Jahitan Blus Antara Pola Leeuw Van Rees dengan Pola m.h Wancik untuk Wanita Bertubuh Gemuk*," pp. 23–27.

[18] A. Maula, 2013, "*Perancangan dan Implementasi Aplikasi Pola Dasar Busana Wanita*".

[19] R. Kirkby, E. Frank, and P. Reutemann, 2007, "WEKA Explorer User Guide for Version 3-5-7".

[20] S. Drazin and M. Montag, "*Decision Tree Analysis Using Weka*," pp. 1–3.