

# KLASIFIKASI PENENTUAN TIM UTAMA OLAHRAGA HOCKEY MENGUNAKAN ALGORITMA C4.5 (STUDI KASUS : HOCKEY KABUPATEN KENDAL)

Budi Utami<sup>1</sup>, Yuniarsi Rahayu<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Jl. Nakula I No 5-11 Semarang 50131, Telp : (024) 351-7261, Fax : (024) 352-0165  
E-mail : amoy.baehaqi@gmail.com<sup>1</sup>, yuniarsi.rahayu@dsn.dinus.ac.id<sup>2</sup>

## Abstrak

Klasifikasi pemain hockey dilakukan untuk mengklasifikasi pemain yang dapat lolos masuk tim utama dan gagal masuk tim utama. Klasifikasi dilakukan menggunakan data mining algoritma C4.5. Data yang digunakan untuk penelitian meliputi data multilevel pemain, data sprint, data tembakan push, data control bola, dan data game pemain. Proses data mining pada data training akan menghasilkan sebuah pohon keputusan atau rule. Metode evaluasi yang dilakukan dalam penelitian ini menggunakan confusion matrik dimana nilai akurasi untuk tiga kali pengujian mengalami kenaikan yaitu untuk prosentase data training dan data testing sebesar 70%:30% menghasilkan nilai akurasi sebesar 70%, prosentase data 80%:20% menghasilkan nilai akurasi sebesar 75%, dan untuk prosentase data 90%:10% menghasilkan nilai akurasi sebesar 80%. Ini membuktikan bahwa semakin besar data training maka semakin besar pula nilai akurasi yang didapat.

**Kata kunci :** Hockey, klasifikasi, C4.5, data mining, confusion matrix

## Abstract

Classification of hockey players performed to classify a player who can pass into the first team and failed to make the first team. Classification is done using data mining algorithms C4.5. Data used for the study includes data multilevel players sprint the data, the data push shot, ball control data, and data player game. The process of data mining on training data will result in a decision tree or a rule. Methods of evaluation made in this study using a confusion matrix in which the value of accuracy to three times the test has increased that to a percentage of the training data and the data of testing by 70%: 30% yield accuracy value of 70%, the percentage of data is 80%: 20% yield value accuracy by 75%, and for the percentage of data is 90%: 10% yield value of 80% accuracy. This proves that the larger the training data the greater the accuracy values were obtained.

**Keywords :** hockey, classification, C4.5, data mining, confusion matrix

## 1. PENDAHULUAN

Hockey merupakan permainan olahraga yang dilakukan oleh pria dan wanita dengan menggunakan alat pemukul (stick) dan bola. Bentuk permainannya hampir sama dengan sepak bola namun pada hockey bola tidak boleh mengenai kaki pemain. Di Jawa Tengah hockey yang populer ada di beberapa kota/kabupaten, salah satunya ada di kabupaten Kendal. Tim hockey kabupaten Kendal lahir pada tahun 2005. Seiring dengan pesatnya

olahraga hockey di kabupaten Kendal terutama di Boja dan sekitarnya, sekarang jumlah anggota tim hockey kabupaten Kendal lebih dari 50 orang. Anggota tersebut terdiri dari pria dan wanita.

Dalam proses seleksi pemain untuk dapat masuk ke tim utama, biasanya tim kepelatihan hanya menggunakan data training match. Padahal dalam proses seleksi pemain seharusnya ada kriteria-kriteria yang telah ditentukan sebagai syarat masuk atau tidaknya seorang

pemain kedalam tim utama.

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistic, matematika, kecerdasan buatan, dan *machine learning* [1]. Salah satu metode data mining adalah klasifikasi. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data [2].

Oleh karena itu, dalam penelitian ini akan menggunakan algoritma C4.5. Algoritma C4.5 merupakan perkembangan dari algoritma ID3 yang mudah dimengerti serta dapat membangun pohon keputusan dengan cepat [3].

## 2. METODE PENELITIAN

### 2.1 Klasifikasi

Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasikan suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasikan spesies berdasarkan karakteristik fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi. Komponen-komponen utama dari proses klasifikasi antara lain [3]:

- 1) Kelas, merupakan variable tidak bebas yang merupakan label dari hasil klasifikasi. Sebagai contoh adalah kelas loyalitas pelanggan, kelas badai atau gempa bumi, dan lain-lain.
- 2) Prediktor, merupakan variable bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi, misalnya merokok, minum-minuman beralkohol, tekanan

darah, status perkawinan, dan sebagainya.

- 3) Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan predictor untuk dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat. Contohnya adalah grup pasien yang telah di-test terhadap serangan jantung, grup pelanggan di suatu supermarket, dan sebagainya.
- 4) Set data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

### 2.2 Algoritma C4.5

Algoritma pohon keputusan yang paling terkenal adalah C4.5. Pada akhir tahun 1970 sampai awal tahun 1980 J.Ross Quinlan, seorang peneliti di bidang machine learning, membuat sebuah algoritma decision tree yang dikenal dengan ID3 (Iterative Dichotomisher). Quinlan kemudian membuat algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan). Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bias dibagi lagi. Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu [3]:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i \quad (1)$$

Keterangan :

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

3. Kemudian hitung nilai gain menggunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S = Himpunan kasus

A = fitur

n = jumlah partisi atribut A

|Si| = proporsi Si terhadap S

|S| = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua record terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
  - a. Semua record dalam simpul N mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam record yang dipartisi lagi.
  - c. Tidak ada record di dalam cabang yang kosong.

### 2.3 Confusion Matrix

*Confusion Matrix* adalah tools yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai actual dan prediksi pada klasifikasi [4].

**Tabel 1** : Klasifikasi dan Prediksi

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a ( <i>true positive-TP</i> )	b ( <i>false negative-FN</i> )
Class = No	c ( <i>false positive-FP</i> )	d ( <i>true negative-TN</i> )

Evaluasi dan validasi hasil dihitung menggunakan rumus akurasi, *precision recall* dan *f-measure* berikut ini [5]:

1. Akurasi

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample *data testing* yang diuji.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2. *Precision*

Menghitung nilai *precision* dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Negative*).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. *Recall*

Sedangkan *recall* dihitung dengan cara membagi data benar yang bernilai positif (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data

salah yang bernilai negatif (*False Negative*).

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

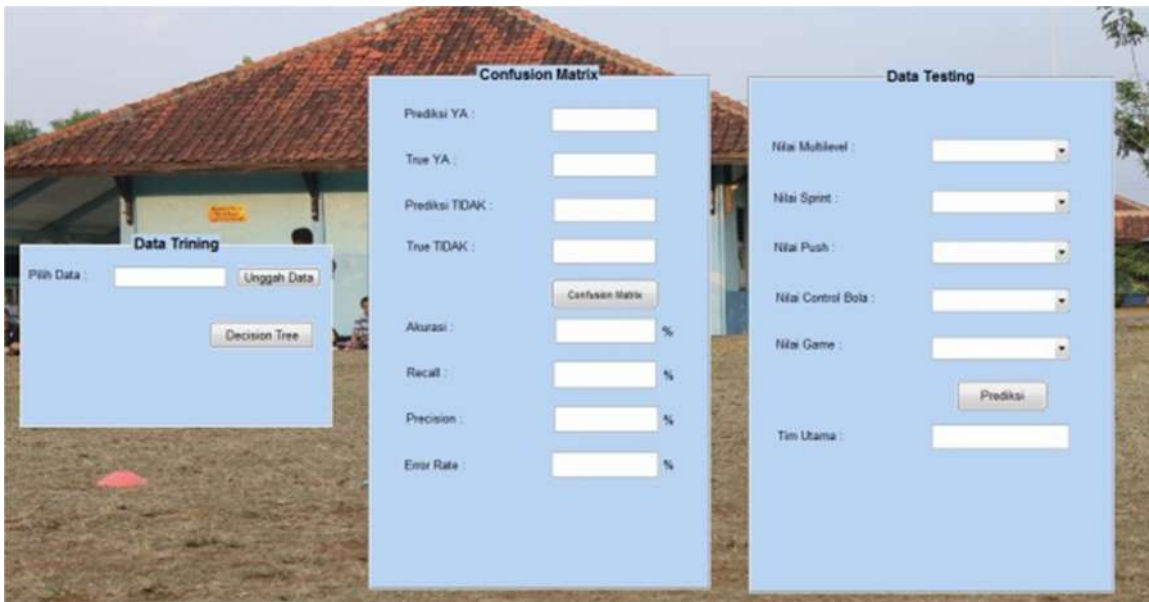
$$F - Measure = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

#### 4. F-Measure

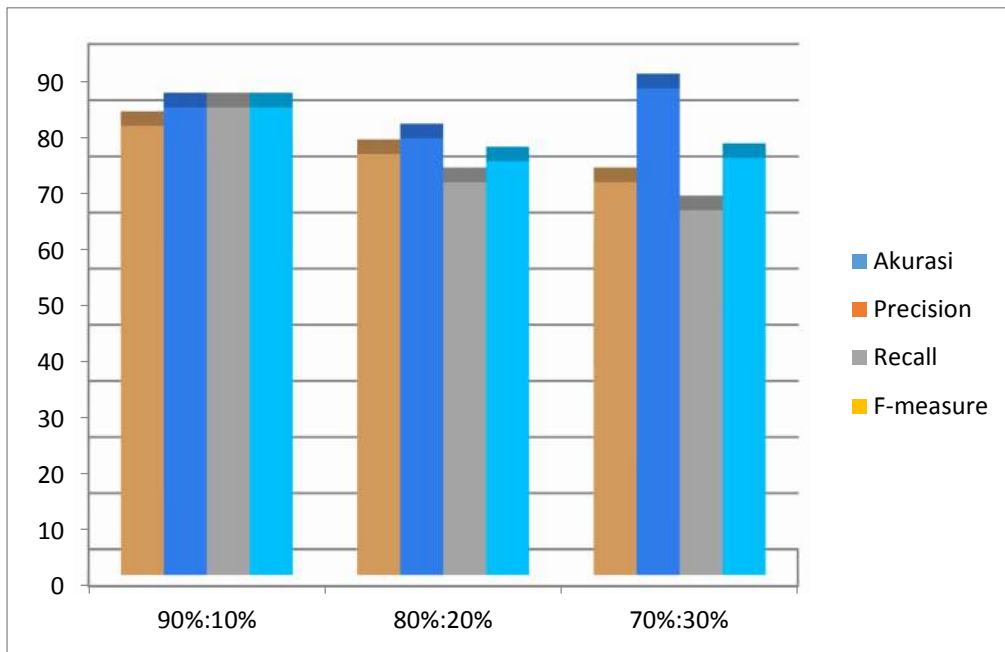
Nilai *F-Measure* didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini, menerapkan algoritma klasifikasi C4.5 yang diimplementasi menggunakan software matlab. Berikut merupakan hasil implementasinya.



Gambar 1. Tampilan Program



Gambar 2. Grafik Hasil Evaluasi

Grafik menunjukkan ada penurunan terhadap nilai akurasi dan *recall*. Sedangkan untuk nilai *precision* mengalami penurunan dari 83.3333% menjadi 77.7778% dan meningkat kembali menjadi 86.6667%. Pada nilai *f-measure* juga mengalami penurunan dari 83.3333% menjadi 73.6842% dan mengalami peningkatan kembali menjadi 74.2857%. Semakin banyak data training maka semakin akurat hasil yang didapat.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian dari permasalahan yang dikembangkan maka dapat disimpulkan bahwa untuk study kasus seleksi pemain hockey dapat memanfaatkan teknik klasifikasi data mining menggunakan algoritma C4.5 sebagai klasifikasi Lolos atau Gagal untuk dapat masuk tim utama hockey kabupaten Kendal. Data yang digunakan adalah data yang didapatkan dari pemain hockey kabupaten Kendal. Dari metode klasifikasi data mining dengan algoritma C4.5 yang dipilih menghasilkan sebuah pohon keputusan dan rules. Dari rules yang dihasilkan maka dilakukan evaluasi terhadap rules menggunakan *confusion matrix*. Dari perhitungan confusion matrik menghasilkan peningkatan nilai akurasi

apabila data semakin banyak maka nilai akurasi semakin tinggi.

#### DAFTAR PUSTAKA

- [1] Primadi Tabrani, *Hockey & Kreativita dalam Olah Raga*. Bandung: ITB, 1985.
- [2] Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques*, 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2006.
- [3] Rahmadya T. H dan Herlawati Prabowo P. W, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains, 2013.
- [4] Dwi Untari, "DATA MINING UNTUK MENGANALISA PREDIKSI MAHASISWA BERPOTENSI NON-AKTIF MENGGUNAKAN METODE DECISION TREE C4.5," Universitas Dian Nuswantoro, Semarang, pdf skripsi 2014.
- [5] Eko Prasetyo, *Mengolah Data Menjadi Informasi Menggunakan Matlab*, Aldo Sahala, Ed. Yogyakarta: ANDI, 2014.