



## **METODE KLASIFIKASI *DECISION TREE* UNTUK MEMPREDIKSI JUARA *ENGLISH PREMIER LEAGUE***

**Syaiful Bahri<sup>1)</sup>, Akhyar Lubis<sup>2)</sup>**

<sup>1,2</sup>Universitas Pembangunan Panca Budi, Medan, Indonesia

[Syiaiful.bahri@dosen.pancabudi.ac.id](mailto:Syiaiful.bahri@dosen.pancabudi.ac.id)

[akhyarlbs@pancabudi.ac.id](mailto:akhyarlbs@pancabudi.ac.id)

### **ABSTRAK**

Pada hasil akhir klasemen liga Inggris setiap musim dapat dianalisis berdasarkan atribut-atribut dari tabel klasemen. Dengan sekumpulan data hasil klasemen setiap akhir musim dapat diklasifikasikan seberapa besar sebuah club tersebut mampu menjadi juara berdasarkan atribut yang paling dominan. Penelitian ini menerapkan metode klasifikasi *decision tree* dengan algoritma *C4.5* menggunakan *software RapidMiner 8.1*. Club yang akan diuji adalah Manchester United. Data set yang digunakan adalah data club dari tahun 1992 hingga tahun 2016, jumlah goal yang diciptakan setiap musimnya, jumlah kemasukan goal tiap musimnya, jumlah *points* diakhir musim dan status juara atau tidak di akhir musim. Hasil penelitian dengan algoritma *C4.5* ini menunjukkan bahwa club tersebut akan mampu menjadi juara di akhir musim jika mampu menciptakan jumlah goal diatas 65 dan *point* yang diperoleh diatas 89 point di setiap akhir musim. Jika jumlah perolehan *points* club tersebut diakhir musim dibawah 65 *points*, maka club tersebut dipastikan tidak pernah juara diakhir musim.

**Kata Kunci:** klasemen, *decision tree*, algoritma *c.4.5*

### **ABSTRACT**

*In the final results of the English League, each season can be analyzed based on the attributes of the table standings. With a set of data, the results of the standings at the end of each season can be classified as how much a club is able to become champions based on the most dominant attributes. This study applies a decision tree classification method with the C4.5 algorithm using RapidMiner 8.1 software. The club to be tested is Manchester United. The data set used is the club data from 1992 to 2016, the number of goals created each season, the number of goals conceded each season, the number of points at the end of the season and the status of champions or not at the end of the season. The results of this research with the C4.5 algorithm show that the club will be able to win at the end of the season if it is able to create a goal number above 65 and points earned above 89 points at the end of each season. If the number of points the club has at the end of the season is below 65 points, then the club is certain to never win at the end of the season*

**Keywords:** standings, *decision tree*, algorithm *c.4.5*



## I. PENDAHULUAN

Hasil pertandingan akan tersimpan dengan tabel klasemen sebagai penentu juara di akhir musim bagi yang menempati peringkat pertama. Dalam hal prediksi melalui *track record* sebuah club dapat diklasifikasikan hasil-hasil klasemen setiap tahunnya. Hasil klasifikasi tersebut dapat diambil prediksi ataupun kesimpulan dalam berbagai macam hal, salah satunya yaitu mengklasifikasikan seberapa besar peluang sebuah club tersebut bisa juara di liga dengan hasil-hasil akhir klasemen di liga beberapa tahun yang lalu. Hasil ini menjadikan satu pedoman bahwa club harus bisa memperoleh nilai dengan jumlah tertentu jika ingin menjadi juara.

*Decision tree* dengan metode C.45 akan membuat klasifikasi dari hasil klasemen liga primer Inggris dari tahun 1992 sampai dengan 2016. Data yang diolah akan mengklasifikasikan jumlah goal per musim, jumlah kemasukan goal per musim, point akhir musim dan status juara atau tidak di akhir musim yang akan dilakukan pengujian dengan sampel club Manchester United. Dengan adanya klasifikasi tersebut, dapat diambil kesimpulan dalam jumlah atribut mana yang paling besar menentukan peluang club tersebut mampu juara liga di setiap musim. Untuk mempermudah dalam klasifikasi data, penelitian ini menggunakan aplikasi *Rapid Miner 8.1*.

## II. METODE PENELITIAN

### 1) *Data Mining*

*Data Mining* merupakan proses menemukan pengetahuan yang menarik, seperti asosiasi, pola, perubahan, struktur signifikan dan anomali, dari sejumlah besar data yang disimpan dalam *database* atau gudang data atau repositori informasi lainnya (Han, 2006). Beberapa teknik yang sering disebut-sebut dalam *literatur DM* antara lain : *clustering*, *classification*, *association rule mining*, *neural network*, dan *genetic algorithm* (Linda wati, 2008).

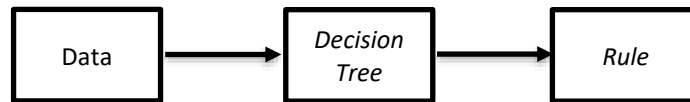
### 2) **Klasifikasi**

Klasifikasi merupakan teknik penambangan data yang memetakan data ke dalam kelompok atau kelas yang telah ditentukan. Ini adalah metode pembelajaran terawasi yang membutuhkan data pelatihan berlabel untuk menghasilkan aturan untuk mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan (Dunham, 2003). Metode klasifikasi mengacu pada pembentukan kelompok data dengan menerapkan algoritma dikenal ke gudang data di bawah pemeriksaan. Metode ini berguna untuk proses bisnis yang membutuhkan informasi kategori seperti pemasaran atau penjualan. Hal ini dapat menggunakan berbagai algoritma seperti sebagai tetangga terdekat, pohon keputusan belajar dan lain-lain. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.



### 3) Decision Tree

*Decision tree* merupakan teknik model prediksi yang dapat digunakan untuk klasifikasi dan prediksi tugas. *Decision tree* menggunakan teknik “membagi dan menaklukkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah (dunham, 2003). Proses pada *decision tree* adalah mengubah bentuk data tabel menjadi sebuah model *tree*. Model *tree* akan menghasilkan *rule* dan disederhanakan (Basuki & Syarif, 2003).



**Gambar 1 Konsep Pohon Keputusan**

### 4) Algoritma C4.5

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, Larose yaitu :

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan 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.

#### Rumus Nilai Entropy

$$Entropy(S) = \sum_{i=1} -p_i * \log_2 p_i$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

$p_i$  : Proporsi dari  $S_i$  terhadap S



3. Menghitung nilai *Gain* menggunakan persamaan.

**Rumus Gain**

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in \text{Value}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

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 dan langkah ke 3 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. Tidak ada *record* di dalam cabang yang kosong.

**5) Preprocessing Data**

Data *cleaning* diterapkan untuk menambah isi atribut yang hilang atau kosong, dan merubah data yang tidak konsisten.

1. Data Transformasi Dalam proses ini, data ditransformasikan ke dalam bentuk yang sesuai untuk proses *data mining*.

2. Data reduksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan sehingga ukuran dari *database* menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses *data mining*, karena akan lebih efisien terhadap data yang lebih kecil. Berikut ini adalah tabel data set yang akan digunakan.

**Tabel 1 Data Set Hasil Preprocessing**

No	Musim	Jumlah Goal Per Musim	Jumlah Kemasukan Goal Per Musim	Point Akhir	Status Juara
1.	2016-17	54	29	69	Tidak
2.	2015-16	49	35	66	Tidak
3.	2014-15	62	37	70	Tidak
4.	2013-14	64	43	64	Tidak
5.	2012-13	86	43	89	Juara
6.	2011-12	89	33	89	Tidak
7.	2010-11	78	37	80	Juara
8.	2009-10	86	28	85	Tidak



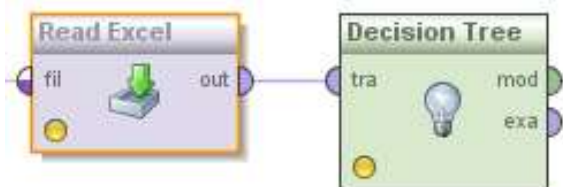
9.	2008–09	68	24	90	Juara
10.	2007–08	80	22	87	Juara
11.	2006–07	83	27	89	Juara
12.	2005–06	72	34	83	Tidak
13.	2004–05	58	26	77	Tidak
14.	2003–04	64	35	75	Tidak
15.	2002–03	74	34	83	Juara
16.	2001–02	87	45	77	Tidak
17.	2000–01	79	31	80	Juara
18.	1999–00	97	45	91	Juara
19.	1998–99	80	37	79	Juara
20.	1997–98	73	26	77	Tidak
21.	1996–97	76	44	75	Juara
No	Musim	Jumlah Goal Per Musim	Jumlah Kemasukan Goal Per Musim	Point Akhir	Status Juara
22.	1995–96	73	35	82	Juara
23.	1994–95	77	28	88	Tidak
24.	1993–94	80	38	92	Juara
25.	1992–93	67	31	84	Juara

### III. HASIL PENELITIAN DAN PEMBAHASAN

#### 1) Menentukan Nilai Atribut

Dalam data sampel tentukan dulu *node* terpilih, yaitu dengan menghitung nilai informasi gain masing-masing atribut untuk menentukan *node* terpilih, gunakan nilai informasi *gain* yang paling besar. Dengan menggunakan persamaan  $-P(+)\log_2P(+)-P(-)\log_2P(-)$ .

Pada tahap ini penelitian akan dilakukan dengan menggunakan *software RapidMiner 8.1* dengan metode *decision tree*. Data yang diolah menggunakan *M.excel 2010*, sehingga di aplikasi *RapidMiner 8.1* menggunakan operator *read excel* yang berfungsi untuk membaca file yang akan diolah. Kemudian masukan operator *Decision Tree* untuk membuat data yang diolah menghasilkan pohon keputusan.



Gambar 2 Proses Awal *RapidMiner*

#### 2) Menentukan Parameter

Kriteria yang digunakan adalah *gain\_ratio* berdasarkan informasi yang diperoleh terhadap informasi intristik yang berfungsi untuk mengurangi bias terhadap atribut multi nilai dengan memperhitungkan jumlah dan ukuran cabang ketika memilih atribut. Pada



pilihan *maximal depth* menunjukkan jumlah maksimal panjang percabangan dari puncak pohon ke cabang-cabangnya.

Parameters

Decision Tree

criteria: gain\_ratio

maximal depth: 20

apply pruning

confidence: 0.25

apply prepruning

minimal gain: 0.1

minimal leaf size: 2

minimal size for split: 4

number of prepruning alternatives: 3

**Gambar 3 Pengaturan Parameter**

### 3) Example Set

*Example set* adalah *data set* yang sudah terupload di aplikasi *RapidMiner 8.1*. Berikut adalah *data set* yang siap untuk diproses dengan aplikasi *RapidMiner 8.1*.

**Tabel 2 Example Set**

No	Status Juara	Musim	Jumlah Goal Per Musim	Jumlah Kemasukan Goal Per Musim	Points
4	No	2013-2014	64	43	54
5	Yes	2012-2013	86	43	89
6	No	2011-2012	89	33	89
7	Yes	2010-2011	78	37	80
8	No	2009-2010	86	28	85
9	Yes	2008-2009	68	24	90
10	Yes	2007-2008	80	22	87
11	Yes	2006-2007	83	27	89
12	No	2005-2006	72	34	83
13	No	2004-2005	58	26	77
14	No	2003-2004	64	35	75
15	Yes	2002-2003	74	34	83
16	No	2001-2002	87	45	77



17	Yes	2000-2001	79	31	80
18	Yes	1999-2000	97	45	91
19	Yes	1998-1999	80	37	79
20	No	1997-1998	73	26	77

#### 4) Description

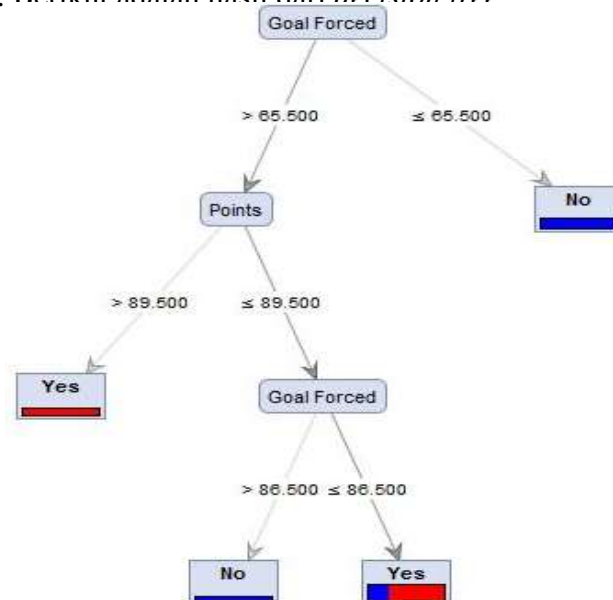
Description menjelaskan hasil dari klasifikasi data set yang diolah. Hasil dari description dapat lihat pada gambar berikut.

```
Tree
Goal Forced > 65.500
| Points > 89.500: Yes {No=0, Yes=3}
| Points ≤ 89.500
| | Goal Forced > 86.500: No {No=2, Yes=0}
| | Goal Forced ≤ 86.500: Yes {No=4, Yes=10}
Goal Forced ≤ 65.500: No {No=6, Yes=0}
```

Gambar 4 Example set read excel

#### 5) Graph View

Graph view menunjukkan hasil dari klasifikasi dengan percabangan yang dapat dihasilkan kesimpulan. Berikut adalah hasil dari decision tree



Gambar 5 Hasil Decision Tree



#### IV. SIMPULAN

##### 1) Kesimpulan

Hasil dari data yang sudah diolah menunjukkan bahwa club *Manchester United* akan menjadi juara jika perolehan pointnya setiap musim 89 point atau >89 point. Jika point di akhir musim dibawah 65 point maka dipastikan club tersebut tidak akan juara dan club tersebut juga mampu mendapatkan gelar juara 7 kali selain dari jumlah point lebih dari 89 point dengan rata-rata goal permusim dibawah 86.

##### 2) Saran

Data set yang akan diolah harus lebih banyak agar hasil lebih optimal dan akurasi lebih tepat.

#### DAFTAR PUSTAKA

- Basuki, Ahmad dan Syarif, Iwan, 2003. "*Decision Tree*". Surabaya: Politeknik Elektronika Negeri
- Dunham, Margareth H., 2003. "*Data Mining Introductory and Advanced Topics*". New Jersey: Prentice Hall.
- Han, J. and Kamber, M., 2006. *Data Mining: Concepts and Techniques*, Elsevier.
- Tjahyono, A. dan Anggara, A. M., 2010. *Sistem Pendukung Keputusan Penerimaan Pegawai Baru pada PT. Kanasritex Semarang, Techno.com, Vol. 9 No.3*
- Kursini, Luthfi. E. T. 2009. *Algoritma Data Mining*. PT Andi Offset.
- Lindawati., 2008. *Data Mining Dengan Teknik Clustering Dalam Pengklasifikasian Data Mahasiswa Studi Kasus Prediksi Lama Studi Mahasiswa Universitas Bina Nusantara, Seminar Nasional Informatika (semnasIF 2008), ISSN : 1979-2328*