

IMPLEMENTASI METODE DATA MINING UNTUK MEMPREDIKSI PERTANDINGAN SEPAK BOLA

Arif Alfa Risal, Abdullah

Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer
Universitas Islam Indragiri (UNISI)
Jl. Propinsi, Parit 1 Tembilahan Hulu, Tembilahan, Riau, Indonesia
Email : aarifalfarisal@gmail.com, abdialam@yahoo.co.id

ABSTRAK

Banyak pengguna kesulitan dalam memprediksi hasil pertandingan sepak bola. Bukan hanya memprediksi hasil pertandingan, pengguna juga kesulitan untuk mengetahui kekuatan klub bola yang akan bertanding dan kesulitan untuk memprediksi klub bola yang di nilainya akan menang. Dengan adanya sistem ini diharapkan untuk bisa membantu pengguna mengatasi masalah tersebut sehingga nantinya pengguna dapat memprediksi hasil pertandingan sepak bola dengan baik, mengetahui kekuatan klub bola yang akan bertanding dan Mengetahui hasil prediksi klub bola yang akan menang. *Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*mechine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara lain. Hasilnya proses metode ini dapat memprediksi pertandingan Sepakbola dengan baik sesuai dengan data – data yang ada, dapat menghemat waktu untuk memprediksi pertandingan tersebut dan mengetahui hasil pertandingan yang telah di prediksinya .

Kata Kunci : Sepakbola, *Data mining*, Pohon keputusan (*decision tree*).

ABSTRACT

Many users difficulty in predicting the outcome of football matches . Not only predict the outcome of a match , the user is also difficult to know the power of football clubs who will compete and difficult to predict which football club in value will win . With this system is expected to help users overcome these problems so that users will be able to predict the results of football matches well , knowing the power of football clubs who will compete and predicted results Knowing ball club will win. Data mining is a process that employs one or more computer learning techniques (a machine learning) to analyze and extract knowledge (knowledge) is another . The result of this method can predict the Football match well in accordance with the data - the data available, it can save time to predict the game and know the results of matches that have been in the predictions.

Keywords: *Football, data mining, decision tree (decision tree).*

1 PENDAHULUAN

Sepakbola adalah olahraga berskala internasional, permainan tersebut membangkitkan luapan keinginan dan emosi yang tidak sama dengan olahraga lainnya. Sepakbola merupakan sesuatu yang umum di antara orang-orang dengan latar belakang dan keturunan yang berbeda-beda, sebuah jembatan yang menghubungkan jejang ekonomi, politik, kebudayaan dan agama. Dikenal sebagai bola kaki hampir di seluruh dunia, sepak bola merupakan olahraga nasional hampir di seluruh negara di Asia, Afrika, Eropa dan Amerika Selatan. Olahraga ini merupakan satu-satunya jenis permainan bola kaki yang dimainkan di olimpiade.

2 TINJAUAN PUSTAKA

2.1 Pekerjaan Dalam *Data Mining*

Pekerjaan yang berkaitan dengan *data mining* dapat dibagi menhadi empat kelompok, yaitu model prediksi (*prediction modelling*), analisis kelompok (*cluster analysis*), analisis asosiasi

(*association analysis*), dan deteksi anomali (*anomaly detection*). Ilustrasi empat pekerjaan tersebut dapat dilihat pada gambar dibawah ini (Prasetyo, 2014).

1. Model Prediksi

Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai terget pada himpunan baru yang didapat. Ada dua jenis model prediksi, yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk variabel target diskret, sedangkan regresi untuk variabel target kontinu (Prasetyo, 2014).

2. Analisis Kelompok

Analisis kelompok melakukan pengelompokkan data-data kedalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data-data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan dengan kelompok tersebut (Prasetyo, 2014).

3. Analisis Asosiasi

Analisis asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya merepresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efisien (Prasetyo, 2014).

4. Deteksi Anomali

Deteksi anomaly berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain. Data-data yang karakteristiknya menyimpang (berbeda) dari data lain disebut outlier. Algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju eror yang rendah. Deteksi anomali dapat diterapkan pada sistem jaringan untuk mengetahui pola data yang memasuki jaringan sehingga penyusupan bisa ditemukan jika pola kerja data yang datang berbeda. Perilaku kondisi cuaca yang mengalami anomali juga dapat dideteksi dengan algoritma ini (Prasetyo, 2014).

2.2 Pohon Keputusan (*Decision Tree*)

Pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam *data mining* dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus. Hal ini karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Akurasi sangat baik asalkan data yang akan diujikannya patokan merupakan data yang akurat (Widodo, 2013).

2.3 Algoritma C4.5

Algoritma pohon keputusan yang 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 Dichotomiser). 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 di mana terdapat simpul yang mendeskripsikan atribut – atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursi mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi (Widodo, 2013).

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Widodo, 2013) yaitu :

- a. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas – kelas tertentu .
- b. 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

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

Keterangan :

- S = himpunan kasus
- n = jumlah partisi S
- P_i = proporsi S_i terhadap S

c. Menghitung nilai gain menggunakan rumus :

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

Keterangan:

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- |S_i| : Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

d. Ulangi langkah ke-2 hingga semua *record* terpartisi .

e. Proses partisi pohon keputusan akan berhenti saat :

- 1) Semua record dalam simpul N mendapat kelas yang sama .
- 2) Tidak ada atribut di dalam *record* yang dipartisi lagi .
- 3) Tidak ada *record* di dalam cabang yang kosong .

3 ANALISA DAN PERANCANGAN SISTEM

3.1 Analisis Data

Semua data sepak bola akan dicari hubungannya dengan data tim bola, hanya beberapa atribut yang kira-kira berguna dan datanya tidak terlalu acak. Karena data yang acak akan membuat proses mining memakan waktu yang lama dan tingkat hubungannya pun rendah. Data sepak bola yang akan dicari hubungannya meliputi proses Kualitas Pemain, Umur, Kualitas Pelatih dan Semangat Tim.

Adapun variabel yang akan diproses mining meliputi :

1) Kualitas Pemain

Pemain melakukan perubahan posisi yang tidak sesuai dengan posisi aslinya untuk membantu serangan atau pertahanan, dan itu salah satu bentuk takti dan strategi dalam sepakbola . Hal ini sah-sah saja dilakukan tetapi yang perlu diperhatikan adalah segera kembali ke posisi semula setelah melakukan improvisasi, karena akan mengacaukan permainan dan formasi pemain yang telah diatur oleh pelatih dan juga dapat menguras tenaga pemain. Pemain memiliki keahlian tersendiri dalam melakukan improvisasi ini hanya beberapa pemain yang sanggup melakukan posisi seperti ini

2) Umur

Umur sangat berpengaruh terhadap permainan sepakbola, karena seorang pemain produktif bermain sepakbola antara umur 23 – 30 tahun. Umur mempengaruhi permainan sepakbola karena umur ini menentukan faktor fisik seseorang yang bermain sepakbola. Jika umur sudah menginjak kepala tiga maka produktifitasnya untuk bermain sepakbola itu menurun dan semakin tinggi usianya maka semakin meredup pula permainan sepakbolanya .

3) Kualitas Pelatih

Pelatih juga ikut menentukan kualitas permainan dari suatu tim sepakbola. Maka ketika ingin membuat prediksi pertandingan sepakbola selain melihat daftar pemain, perhatikan pula siapa pelatihnya. Karena pelatih juga punya peran yang tidak sedikit dalam suatu kemenangan di pertandingan atau kompetisi.

4) Semangat Tim

Ketika sebuah tim melakukan pertandingan di tempat dia sendiri atau disebut dengan laga kandang, maka kemungkinan untuk menang juga lebih besar. Ini sudah terbukti banyak. Apalagi bila kualitas pemain dan pelatihnya seimbang. Karena jika pertandingan itu dilakukan di kandang sendiri maka pendukung tim akan mejadi penyemangat tim utuk memenangkan pertandingan. Penonton

sepakbola biasanya di sebut pemain ke duabelas dalam suatu pertandingan sepakbola. Apalagi bila jumlahnya sangat banyak dan punya rasa fanatik yang tinggi pula. Hal ini bisa memberi tekanan yang cukup tinggi bagi mental lawan. Prediksi pertandingan sepakbola juga bisa ditentukan dengan melihat faktor ini.

3.2 Analisis Algoritma C4.5

Pada Tabel 3.1 di bawah ini implementasi perhitungan kasus Algoritma C4.5 untuk menentukan kemenangan Sepakbola berdasarkan 3 klasifikasi yaitu kualitas Pelatih, umur dan kualitas pemain. Klasifikasi ini dijadikan atribut awal dalam implementasi algoritma C4.5 dan atribut tujuannya adalah semangat tim dengan klasifikasi dari atribut semangat tim adalah semangat dan tidak semangat.

Tabel 3.1 Tabel Data untuk Klasifikasi Klub Sepakbola

NAMA KLUB	KUALITAS PEMAIN	UMUR	KUALITAS PELATIH	SEMANGAT TIM
AS ROMA	81	26	90	SEMANGAT
ATALANTA	74	27	75	SEMANGAT
BOLOGNA	75	25	80	TIDAK SEMANGAT
CARPI	75	29	75	TIDAK SEMANGAT
CHIEVO	76	26	70	SEMANGAT
EMPOLI	76	25	75	TIDAK SEMANGAT
FIorentina	76	25	80	TIDAK SEMANGAT
FROSINONE	75	26	70	TIDAK SEMANGAT
GENOA	75	24	85	TIDAK SEMANGAT
INTER	81	26	85	SEMANGAT
JUVENTUS	81	27	90	SEMANGAT
MILAN	79	27	80	TIDAK SEMANGAT
NAPOLI	79	26	80	SEMANGAT
PALERMO	75	25	75	SEMANGAT
SAMPDORIA	75	25	80	TIDAK SEMANGAT
SASSUOLO	75	27	75	TIDAK SEMANGAT
SS LAZIO	76	26	85	SEMANGAT
TURIN	76	26	80	SEMANGAT
UDINESE	76	26	80	SEMANGAT
VERONA	75	27	80	TIDAK SEMANGAT

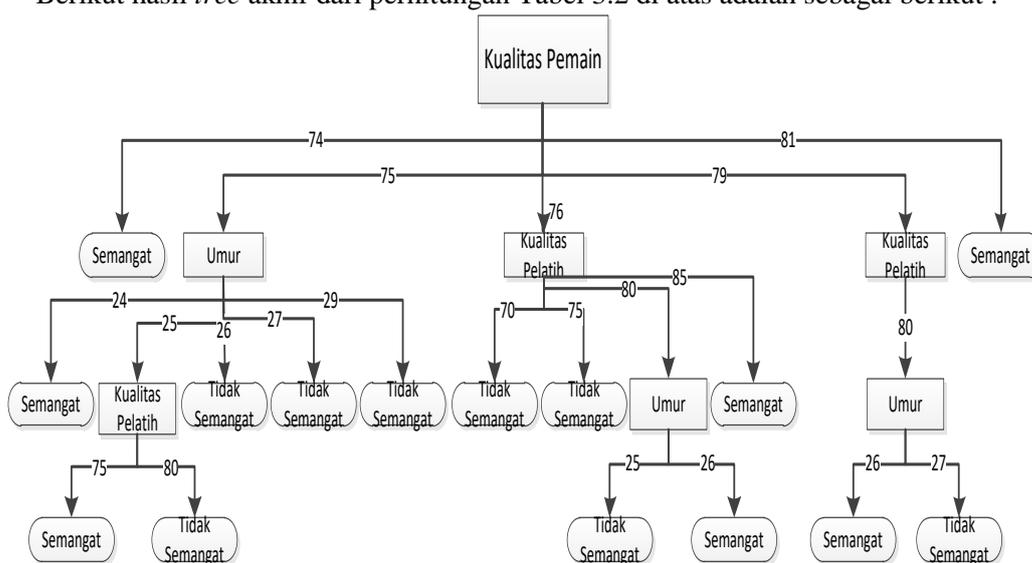
Tabel 3.2 berikut ini adalah penjelasan lebih terperinci mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 menyelesaikan permasalahan dalam skripsi. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik. Salah satu proses kalkulasi dari *entropy* adalah proses kalkulasi nilai *entropy* kelancaran.

Tabel 3.2 Nilai Entropy dan Gain

SIMPUL		DATA	SEMANGAT	TIDAK SEMANGAT	ENTROPY	GAIN
AKAR	TOTAL	20	10	10	1	
	KUALITAS PEMAIN					0.2475
	74	1	1	0	0	
	75	8	2	6	0.8812	
	76	6	3	3	1	

	79	2	1	1	1	
	81	3	3	0	0	
UMUR						0.1327
	24	1	1	0	0	
	25	5	2	3	0,9710	
	26	8	5	3	0,9544	
	27	5	2	3	0,9710	
	29	1	0	1	0	
KUALITAS PELATIH						0.1377
	70	2	1	1	1	
	75	5	2	3	0,9710	
	80	8	3	5	0,9544	
	85	3	2	1	0,9183	
	90	2	2	0	0	

Berikut hasil *tree* akhir dari perhitungan Tabel 3.2 di atas adalah sebagai berikut :



Gambar 3.1 Cabang perhitungan kasus algoritma

4 IMPLEMENTASI DAN PEMBAHASAN

4.1 Halaman Utama

Halaman utama merupakan tampilan masternya sistem ini, pada ini terdapat beberapa menu, seperti menu laporan, peserta pertandingan, klub pemain, pemain, pelatih, klub dan user, dari menu tersebut memiliki beberapa sub menu .

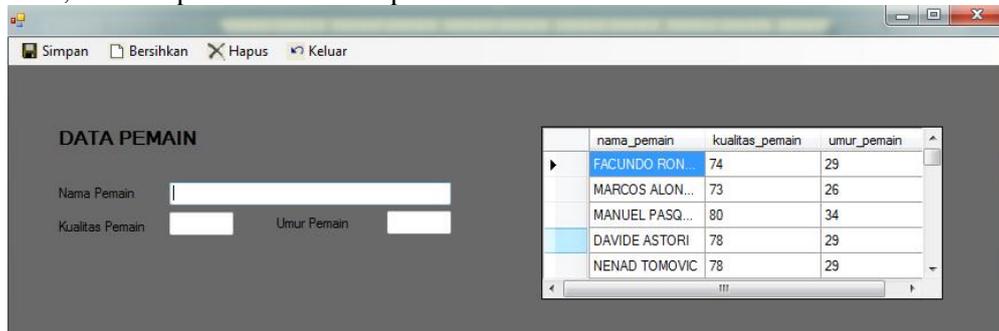


Gambar 4.1 Tampilan Halaman Utama

Pada Gambar 4.1 merupakan gambar tampilan halaman utama, halaman yang memiliki sub menu diantaranya, menu laporan yang didalamnya terdiri dari laporan daftar klub, laporan daftar pelatih, laporan daftar pemain, laporan daftar pemain klub dan data peserta.

4.2 Data Pemain

Tampilan pemain merupakan menu pemain yang berisikan semua pemain pada suatu klub yang akan mengikuti pertandingan yang akan berlangsung, yang mana didalam tampilan tersebut terdapat nama pemain, kualitas pemain dan umur pemain .

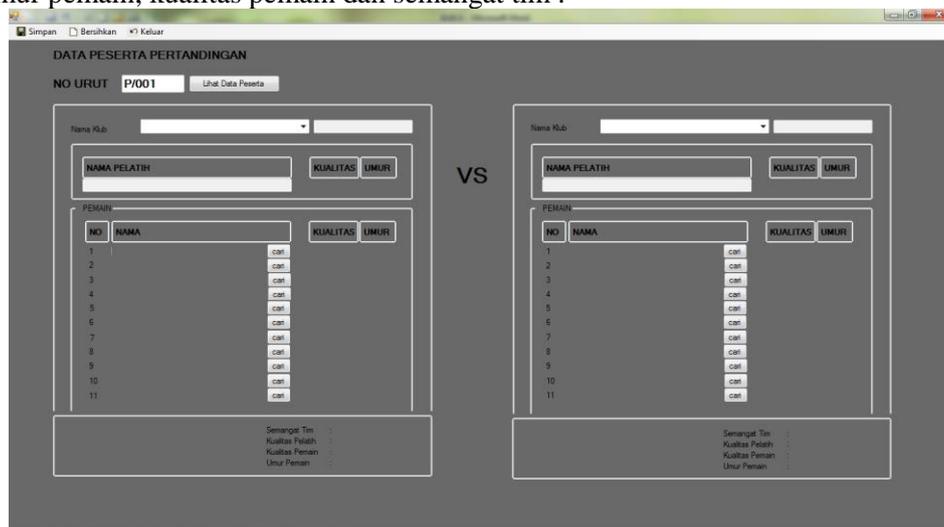


Gambar 4.2 Tampilan Data Pemain

Pada Gambar 4.2 merupakan tampilan menu untuk melakukan penginputan nama pemain, kualitas pemain dan umur pemain, data pemain tersebut harus sesuai dengan data yang ada dalam database yang ada. Kemudian klik tombol simpan.

4.3 Tampilan Peserta Pertandingan

Tampilan peserta pertandingan merupakan menu peserta yang akan bertanding pada laga tersebut, yang mana didalam tampilan tersebut terdapat nama klub, nama pelatih, kualitas pelatih, nama pemain, umur pemain, kualitas pemain dan semangat tim .



Gambar 4.3 Tampilan Peserta Pertandingan

Pada Gambar 4.3 merupakan tampilan peserta pertandingan, yang mana pada form ini akan memprediksi hasil pertandingan sepak bola apakah klub itu akan menang seri atau kalah.

5 PENUTUP

5.1 Kesimpulan

Sistem ini membantu bagi penggunanya. Adapun keuntungan yang dapat diambil dengan adanya sistem yang baru ini adalah memprediksi pertandingan sepakbola lebih akurat dengan data – data yang ada, dapat lebih menghemat waktu dalam memprediksi pertandingan sepakbola yang akan bertanding dan Mengetahui hasil pertandingan sepakbola tersebut .

5.2 Saran

Penulis memberikan saran-saran agar sistem ini bisa bermanfaat bagi pengguna sehingga dengan adanya sistem ini pengguna diharapkan tidak menyalahgunakan program ini untuk kepentingan yang tidak – tidak, seiring berkembangnya sepakbola modern maka akan terjadi pula peningkatan sistem sehingga menyebabkan sistem yang ada akan selalu di perbaharui sesuai dengan data – data yang ada maka dari itu pada saatnya nanti perlu dilakukan evaluasi dan pengembangan sistem lebih lanjut dan pengguna perlu melakukan peninjauan terhadap sistem yang diusulkan agar dapat mempertimbangkan baik dan buruknya sistem sehingga dapat berfungsi sebagaimana yang diharapkan.

REFERENSI

- Budi, Ronald. *Visual Basic*. Malang: Skripta Media Creative
- Fathansyah, 2012. *Basis Data*. Bandung: Informatika Bandung
- Herawati A. Fajar. *Data Mining*. Yogyakarta: Andi
- Kadir, Abdul, 2003. *Pengenalan Sistem Informasi*. Yogyakarta: Andi
- Kadir, Abdul, 2009. *Dasar Perancangan dan Implementasi Database Relasional*. Yogyakarta: Andi
- Ladjamudin, Al-bahra, 2005. *Analisa dan Desain Sistem Informasi*. Tanggung: Graha Ilmu
- Luxbacher A. Joseph. *Sepak Bola*. Jakarta: Rajagrafindo Persada
- Prasetyo, Eko, 2014. *Mengolah Data Menjadi Informasi Menggunakan Matlab*: Yogyakarta: Andi