

## **BAB 2**

### **TINJAUAN PUSTAKA**

Pada bab ini akan dibahas mengenai hasil dari penelitian terdahulu dan landasan teori

#### **2.1 Penelitian Terdahulu**

Pada sub-bab ini, berisi simpulan dan ulasan dari hasil – hasil penelitian terdahulu yang berkaitan dengan penelitian yang akan dilakukan oleh penulis pada penelitian kali ini, antara lain:

##### **2.1.1 Penelitian Pertama**

Penelitian yang berjudul “ Seleksi Penyerang Utama Menggunakan K-Means Clustering dan Sistem Pendukung Keputusan Metode Topsis “[1]. Menggunakan metode K-Means untuk memilih penyerang utama pada kesebelasan tim nasional Perancis, terdapat 13 data yang kemudian dikelompokkan berdasarkan jumlah penampilan dan gol yang dicetak dari setiap pertandingan di klubnya masing – masing. Penulis membuat dua parameter sebagai acuan untuk pengelompokan sehingga setiap data akan dimasukan berdasarkan jarak terdekat dari setiap parameter yang ada. Setelah didapatkan klaster dari seluruh pemain kemudian dilakukan penghitungan lagi rata – rata nilai dari anggota yang ada dalam klaster untuk menentukan titik pusat yang baru. Setelah itu dilakukan perhitungan ulang untuk memastikan tidak ada perubahan pada titik tengah setiap klaster yang ada. penulis melakukan normasilasi terbobot dengan menggunakan kriteria yang telah ditentukan kepada setiap alternative yang ada, tahapan

berikutnya dilakukan penentuan terhadap matrik solusi ideal positif ataupun negatif ataupun tahap terakhir ialah mencari nilai preferensi untuk setiap alternatif yang ada.

### **2.1.2 Penelitian Kedua**

Penelitian yang berjudul “Sistem Penunjang Keputusan Dalam Pemilihan Pemain Untuk Posisi Tertentu Dalam Sepak Bola”[3]. Pada penelitian ini penulis ingin melakukan analisa dalam pemilihan pemain pada posisi tertentu. Penelitian ini bertujuan untuk membantu pelatih dalam melakukan pemilihan pemain agar tidak terjadi kesalahan serta tidak terjadi tindakan pemilihan yang kurang obyektif pada setiap pemain. Dalam penelitian ini penulis menggunakan metode SHP (*Simple Adaptive Weighting*), tahapan yang dilakukan adalah penentuan bobot setiap kriteria dari suatu posisi yang diinginkan pelatih, menentukan *cost* dan *benefit* dari seluruh kriteria yang ada, penormalisasian data, penghitungan bobot setiap kriteria, melakukan perangkingan dari seluruh alternatif yang telah dilakukan pembobotan. Untuk hasil yang akan dicapai penulis akan membuat sebuah sistem informasi yang difokuskan kepada pelatih dan asisten pelatih agar dapat mengelola data pemain, melihat statistik pemain dalam pertandingan, melihat hasil rekomendasi pemain serta melakukan pengelolaan terhadap posisi pemain. Untuk asisten pelatih pada sistem mereka dapat menentukan posisi yang dicari, menentukan kriteria permainan, menentukan kriteria setiap posisi yang akan dicari, mengelola data pemain yang mengikuti seleksi, melakukan seleksi dari seluruh pemain yang telah melakukan proses seleksi.

### 2.1.3 Penelitian Ketiga

Penelitian yang berjudul “*A Methodology For Football Players Selection Problem Based On Multi-Measurements Criteria Analysis*”[4]. Pada penelitian ini penulis ingin melakukan pemilihan pemain berdasarkan *Multi-Criteria Decision Making (MCDM)*, data yang akan diolah berjumlah 24 pemain yang dibagi menjadi 4 grup berisi 6 pemain untuk setiap grupnya. Pemain yang diuji memilih jenjang usia U17 yang diteliti berdasarkan 12 tahapan uji meliputi 3 tes antropometrik yaitu usia, tinggi badan, dan berat badan, 5 tes kebugaran yaitu lompat tinggi, yoyo, lari cepat 10 meter dan *sprint*. Selain itu ada 4 tes keterampilan yaitu *Juggling, Short Passing, Dribbling, Long Passing, Dribbling, Shooting*. Selanjutnya akan dilakukan perankingan pada setiap pemain berdasarkan hasil dari rangkaian tes yang telah dilakukan menggunakan metode TOPSIS. Penelitian ini memanfaatkan nilai rata-rata serta standar deviasi untuk memastikan setiap grup pemain melalui proses yang baik.

### 2.2 Landasan Teoris

Adapun landasan teori serta istilah-istilah yang digunakan oleh Penulis dalam penelitian ini antara lain:

#### 2.2.1 Liga Inggris

Liga Inggris menjadi salah satu liga yang sangat menarik perhatian penonton sepak bola. Pada setiap musimnya selalu tercipta banyak gol disetiap pertandingannya, setiap klub mempunyai strategi tersendiri untuk mengoptimalkan performa dari setiap penyerang yang mereka miliki. Dalam

beberapa tahun terakhir terdapat beberapa pemain yang menjadi kandidat penyerang terbaik pada liga Inggris diantaranya, Sadio Mane, Sergio Aguero, Harry Kane, dan Eden Hazard. Mereka layak disebut sebagai penyerang terbaik karena kontribusi gol mereka untuk klub masing-masing. Rata-rata penyerang terbaik yang dimiliki oleh setiap klub papan atas dapat mencetak 15 gol atau lebih setiap musimnya. Penyerang yang baik memiliki tingkat rasio gol disetiap pertandingan yang tinggi dan juga tingkat konversi peluang yang baik, selain itu penyerang yang baik juga dapat berperan besar membantu terciptanya gol bagi pemain yang lain.

### **2.2.2 Pengertian Data Mining**

Data Mining adalah tahapan yang menggunakan ilmu matematika, ilmu statistika serta kecerdasan buatan atau *machine learning* untuk melakukan ekstraksi data serta melakukan penidentifikasi terhadap informasi yang terkandung dalam data yang diolah. Dalam data mining terdapat beberapa istilah yang biasa digunakan yaitu KDD (*Knowledge discovery in data base*), analisa pola data atau (*Data Analysis*), ekstraksi pengetahuan (*Knowledge Extraction*), Kecerdasan bisnis (*Business Intellegence*), *data archeology*, *data dregging*[5].

Data mining memiliki berbagai fungsi yang dapat meningkatkan dan membantu dalam pengolahan data guna mendapatkan informasi baru untuk pengguna.

1. **Prediction** berguna untuk menemukan pola dari data dengan menggunakan variabel yang diketahui sebagai acuan untuk mendapatkan variabel lain yang tidak diketahui nilai atau jenisnya.
2. **Description** berguna untuk mendapatkan suatu ciri khas penting dari data dalam suatu basis data.
3. **Classification** Proses yang dilakukan untuk mengidentifikasi model atau fungsi dalam menggabungkan *class* atau konsep dari suatu data.
4. **Assosiaciton** digunakan untuk mendapatkan suatu relasi atau hubungan antar variabel atau nilai yang terpadat pada sejumlah kumpulan data.

Proses pada data mining secara umum dilakukan dari seleksi data dari sumber yang digunakan lalu tahap *preprocessing* untuk pembersihan dan penyesuaian data dan yang terakhir adalah tahap evaluasi sehingga menghasilkan output sesuai yang diharapkan.

1. **Data Selection** adalah proses pengumpulan data dari sumber yang ditentukan yang kemudian akan digunakan untuk penelitian. Data ini dapat berupa apa saja misal gambar, angka, suara ataupun huruf.
2. **Preprocessing** adalah proses tahapan awal untuk pengolahan data yang telah didapatkan, pada tahap ini akan dilakukan pembersihan data ganda, pengecekan konsistensi data, serta melakukan perbaikan terhadap *data record* yang hilang.
3. **Transformation** tahap ini dilakukan penyesuaian dari data yang telah melewati proses *preprocessing* menjadi data yang sesuai dengan

kebutuhan penulis. Proses ini umumnya dipengaruhi oleh jenis dan pola data yang dimiliki.

4. **Data Mining** adalah tahapan inti dari proses pengolahan data. Data yang ada akan diolah menggunakan metode dan algoritma tertentu hingga menghasilkan ekstraksi informasi yang dibutuhkan. Pemilihan metode atau algoritma tergantung dari karakteristik data serta tujuan yang akan dicapai.
5. **Interpretation** tahapan ini merupakan proses terakhir dari rangkaian yang ada. Setelah data diolah dengan metode dan algoritma tertentu pada tahap ini akan disajikan atau ditampilkan sesuai dengan kebutuhan pengguna. Tampilan informasi tentunya menggunakan bahasa serta bentuk yang mudah dimengerti setiap orang agar informasi yang diperoleh dapat disampaikan dengan baik.

### 2.2.3 Pengertian K-Means

K-Means merupakan salah satu algoritma atau metode dalam data mining yang sangat populer untuk digunakan pada dataset dengan jumlah yang besar[6]. Secara garis besar K-Means digunakan untuk membagi atau mempartisi data dalam sebuah cluster[7]. Pengelompokan dilakukan berdasarkan karakteristik data yang sama lalu untuk data yang memiliki karakteristik yang berbeda akan dikelompokkan pada tempat yang lain. Metode K-Means bertugas mengelompokkan data ( $N$ ) pada kluster ( $K$ ) dengan nilai rata-rata terdekat[8], selain itu *K-Means* juga melakukan pemodelan tanpa supervisi (*unsupervised*). Sehingga tujuan utama metode ini adalah memperkecil variasi data yang ada pada suatu setiap cluster dan memaksimalkan variasi pada cluster yang lain. Saat proses iterasi

terjadi pengelompokan kembali data-data yang telah dikelompokan sebelumnya, yaitu secara tegas (*Hard K-Means*) dan menggunakan logika fuzzy atau *Fuzzy K-Means*[9]

#### **2.2.4 Pengertian TOPSIS**

Topsis (*Technique for Order of Preference by Similarity to Ideal Solution*) adalah salah satu metode dalam pengambilan keputusan yang diperkenalkan oleh Yoon dan Hwang pada tahun 1981. Metode pengambilan keputusan berfungsi sebagai alat bantu bagi pengambil keputusan untuk memperluas kemampuan pengambil keputusan tanpa menggantikan penilaian dari individu yang mengambil keputusan[10]. Penggunaan metode ini banyak diterapkan pada pengambilan keputusan yang memiliki kriteria banyak atau multikriteria. Metode ini memiliki konsep dimana penentuan alternatif terbaik dilakukan dengan perhitungan dari jarak terpendek solusi ideal positif dan jarak terjauh untuk solusi ideal negatif, pada metode TOPSIS pembobotan pada setiap indikator merupakan tahapan yang krusial[11], banyaknya faktor yang menjadi pertimbangan membuat proses pengambilan keputusan menjadi agak sulit. Jenis permasalahan tersebut biasanya dinamakan MCDM (*Multiple Criteria Decision Making*), Topsis banyak digunakan karena metode ini sangat sederhana dan mudah untuk dilakukan, dengan konsep yang sederhana serta proses dan langkah – langkah yang mudah dipahami oleh pengguna.

### 2.2.5 Pengertian FDR

FDR (*Fisher's discriminant ratio*) adalah salah satu metode analisa statistic untuk menentukan variabel yang terkait dan variabel tidak terkait berupa variabel kontinu atau binary[12]. Metode ini pertama kali dicetuskan oleh Sir Ronald Fisher pada tahun 1936. FDR pada umumnya digunakan untuk menentukan besarnya perbedaan pada setiap fitur pada dua kelas berbeda. Hasil dari proses FDR adalah untuk fitur yang mempunyai perbedaan yang besar pada rata – rata dari kelas dan varian kecil dari tiap kelas maka nilai FDR yang didapatkan akan tinggi, jika diantara fitur terdapat perbedaan absolut rata – rata yang sama maka fitur dengan varian terkecil akan mendapatkan nilai FDR yang lebih tinggi[13]. Pada penelitian ini metode FDR digunakan untuk menseleksi fitur yang paling ideal untuk digunakan pada proses selanjutnya.