BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Untuk mempersiapkan pemain muda yang memiliki potensi di bidang sepak bola, maka tempat pembinaan usia dini seperti sekolah sepak bola (SSB) perlu menerapkan pelatihan yang tepat bagi setiap pemain. Agar dapat menerapkan metode pelatihan yang tepat maka pelatih perlu mengetahui kemampuan dari setiap pemain yaitu dengan memanfaatkan data kemampuan pemain binaan kemudian dilakukan pengelompokkan berdasarkan kemiripan karakteristik pemain hingga dihasilkan kelompok-kelompok dengan karakteristik tertentu yang dapat disesuaikan dengan metode pelatihan yang tepat bagi masing-masing kelompok. Hal tersebut dapat dilakukan dengan menggunakan teknik *clustering*.

Clustering merupakan metode *data mining* yang digunakan untuk mencari dan mengelompokkan data berdasarkan kemiripan karakteristik antar data, setiap anggota dalam suatu kelompok akan berisi data yang semirip mungkin dan memiliki perbedaan karakteristik dengan kelompok lain. Hasil *cluster* pada *clustering* dapat dikatakan baik jika semakin mirip karakteristik data pada suatu *cluster* dan semakin berbeda dengan *cluster* lain. Agar dapat menghasilkan hasil *clustering* terbaik maka akan sangat bergantung pada data dan algoritma yang digunakan karena setiap data dan algoritma memiliki karakteristiknya masing-masing. Sehingga data dengan karakteristik tertentu akan menghasilkan hasil *clustering* terbaik jika diolah dengan menggunakan algoritma yang memiliki kesamaan karakteristik dengan data. Untuk mengetahui baik atau tidaknya hasil *cluster* yang terbentuk, maka diperlukan pengukuran terhadap hasil *cluster*. Salah satu pengukuran yang dapat digunakan untuk mengetahui kualitas hasil *cluster* adalah Davies-Bouldin Index (DBI). Hasil *cluster* pada pengukuran DBI dapat dikatakan baik jika nilai yang dihasilkan semakin mendekati 0.

Algoritma yang terdapat dalam *clustering* terbagi menjadi dua yaitu hierarki dan partisi. *Clustering* berbasis hierarki memulai proses dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon yang memperlihatkan ada hierarki yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. *Clustering* berbasis hierarki tepat jika diimplementasikan pada data yang belum diketahui jumlah *cluster*nya.

Sedangkan *clustering* berbasis partisi memulai proses dengan menentukan terlebih dahulu dugaan jumlah *cluster* yang diinginkan, selanjutnya menentukan titik pusat awal dan memulai proses *clustering* dengan mengelompokkan titik data berdasarkan kemiripan dan ketidakmiripan. *Clustering* berbasis partisi tepat jika diimplementasikan pada data yang sudah diketahui dugaan jumlah *cluster* yang dapat dibentuk. Sehingga kasus pengelompokkan metode pelatihan sepak bola lebih tepat diimplementasikan pada *clustering* berbasis partisi karena data kemampuan pemain pada kasus tersebut sudah dapat diduga jumlah cluster yang akan dibentuk.

Penentuan dugaan jumlah *cluster* pada partisi dapat disesuaikan dengan kebutuhan pelatih. Akan tetapi, dugaan jumlah *cluster* yang ditentukan tidak dapat dipastikan merupakan jumlah *cluster* terbaik. Hal tersebut merupakan kekurangan dari algoritma partisi. Selain itu *clustering* berbasis partisi juga sensitif terhadap penentuan titik pusat awal (Ali, Arna, & Yuliana, 2005) seperti pada K-Means. Sehingga menyebabkan hasil *clustering* sulit mencapai *global optimum*, akan tetapi terjebak pada kondisi *local minimum*. *Local minimum* merupakan suatu keadaan hasil *clustering* yang terbentuk tidak menjamin hasil yang terbaik, sedangkan *global optimum* merupakan keadaan sebaliknya dari *local minimum*, hasil *clustering* yang terbentuk sudah mencapai hasil yang terbaik.

Berbagai penelitian telah dilakukan untuk menutupi kelemahan pada algoritma *clustering*, seperti pada penelitian yang dilakukan oleh Madu Yedla, dkk (Madu, Srinivasa, & Pathakota, 2010), yang mengusulkan algoritma

Enhanced K-Means. Enhanced K-Means merupakan pengembangan dari algoritma K-Means dengan penambahan algoritma pada penentuan titik pusat awal, sehingga mampu mencari dan menentukan titik pusat awal terbaik. Algoritma yang diusulkan secara efisien dapat menempatkan setiap titik data ke *cluster* yang sesuai. Algoritma Enhanced K-Means terbukti mampu meningkatkan hasil clustering. Akan tetapi algoritma ini masih mengadopsi sensitifitas pada algoritma *clustering*, yaitu perlu menduga jumlah *cluster*. Penentuan dugaan jumlah *cluster* pada algoritma Enhanced K-Means ditentukan oleh pengguna atau secara acak, sehingga baik atau tidaknya hasil *clustering* yang terbentuk bergantung pada dugaan jumlah *cluster* yang telah ditentukan sebelumnya, yang berarti masih memungkinkan terdapat hasil *clustering* terbaik pada jumlah *cluster* yang berbeda. Penelitian lain juga dilakukan untuk memperbaiki sensitifitas pada penentuan dugaan jumlah *cluster* pada *clustering* berbasis partisi seperti yang dilakukan oleh Ahameed, dkk (Ahameed & Hareesha, 2012). Mengusulkan algoritma Dynamic Cluster yang merupakan pengembangan dari algoritma K-Means, pada penelitian ini algoritma Dynamic Cluster dapat menutupi kekurangan pada K-Means yaitu sensitif terhadap penentuan dugaan jumlah *cluster* sehingga algoritma Dynamic Cluster mampu menentukan jumlah *cluster* yang tepat (Widiarina & Romi, 2015). Algoritma Dynamic Cluster bekerja setelah proses K-Means telah selesai, kemudian algoritma Dynamic Cluster akan menghitung jarak antar setiap titik pusat (inter) dan jarak setiap titik data terhadap titik pusat pada cluster yang sama (intra) dari proses sebelumnya, jika jarak intra baru lebih kecil dari jarak intra lama dan jarak inter baru lebih besar dari jarak inter lama maka algoritma Dynamic Cluster akan menambah satu pada jumlah *cluster* sebelumnya, proses ini akan dilakukan berulang hingga algoritma Dynamic *Cluster* menemukan jarak intra yang sekecil mungkin dan jarak inter yang sebesar mungkin.

Berdasarkan hal yang telah diuraikan sebelumnya, maka diketahui bahwa algoritma Enhanced K-Means dan Dynamic Cluster memiliki kelebihan dan kekurangannya masing-masing. Enhanced K-Means memiliki kelebihan yaitu menambahkan algoritma baru untuk menentukan titik pusat awal akan tetapi

memiliki kekurangan yaitu tidak dapat menduga jumlah cluster terbaik. Sedangkan Dynamic Cluster memiliki kelebihan yaitu menambahkan algoritma baru yang dapat menentukan dugaan jumlah *cluster* terbaik. Maka pada penelitian ini akan dilakukan penggabungan algoritma Enhanced K-Means dan Dynamic Cluster yang dapat memperbaiki sensitifitas penentuan dugaan awal jumlah *cluster* dan titik pusat awal sehingga dapat menghasilkan hasil *clustering* terbaik pada kasus pengelompokkan metode pelatihan sepak bola.

1.2 Identifikasi Masalah

Permasalahan yang akan dikaji pada penelitian ini adalah bagaimana cara menggabungkan algoritma Enhanced K-Means dan Dynamic Cluster agar menghasilkan *cluster* yang memiliki nilai DBI terbaik pada data kemampuan pemain binaan. Algoritma Enhanced K-Means memiliki kelebihan yaitu mampu menentukan titik pusat awal terbaik, sedangkan algoritma Dynamic Cluster memiliki kelebihan yaitu dapat menduga jumlah cluster terbaik pada suatu *dataset*. Dari penggabungan algoritma Enhanced K-Means dan Dynamic Cluster diharapkan dapat menghasilkan cluster dengan nilai DBI terbaik pada data kemampuan pemain binaan. Namun pada penelitian sebelumnya yaitu Enhanced K-Means (EKM) (Madu, Srinivasa, & Pathakota, 2010) dan Dynamic Cluster (DKM) (Ahameed & Hareesha, 2012) tidak dijelaskan mengenai karakteristik algoritma pada beragam karakteristik data.

Berdasarkan permasalahan tersebut, maka pada penelitian ini akan dilakukan penggabungan kedua algoritma kemudian menganalisis karakteristik algoritma ketika diimplementasikan pada data model buatan yang telah diketahui dan ditentukan jumlah data, jumlah atribut dan jumlah cluster serta diimplementasikan pada data UCI yang umum digunakan pada penelitian di bidang *data mining*. Setelah diketahui karakteristik dari setiap algoritma selanjutnya adalah mengimplementasikan algoritma pada data kemampuan pemain binaan. Sehingga dapat diketahui algoritma yang paling tepat untuk menyelesaikan kasus pengelompokkan metode pelatihan sepak bola.

1.3 Premis

Berdasarkan permasalahan yang telah diuraikan sebelumnya, maka dapat diformulasikan premis-premis sebagai berikut:

1. Algoritma Enhanced K-Means dapat menentukan titik pusat awal dengan mencari titik pusat terbaik sehingga hasil *clustering* pada Enhanced K-Means tidak akan terjebak pada kondisi *local minima*, namun berada pada kondisi *global optimum*. Akan tetapi algoritma Enhanced K-Means perlu mengetahui jumlah *cluster* yang akan dibentuk.
2. Algoritma Dynamic Cluster dapat menduga jumlah *cluster* terbaik.

1.4 Hipotesis

Penggabungan algoritma Enhanced K-Means dan Dynamic Cluster dapat menghasilkan *cluster* yang memiliki nilai DBI terbaik pada kasus pengelompokkan metode pelatihan sepak bola.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menggabungkan kedua algoritma antara algoritma Enhanced K-Means dengan Dynamic Cluster untuk menghasilkan *cluster* dengan nilai DBI terbaik.
2. Mengetahui karakteristik dari penggabungan algoritma Enhanced K-Means dan Dynamic Cluster.
3. Mengetahui algoritma yang tepat untuk menyelesaikan kasus pengelompokkan metode pelatihan sepak bola.

1.6 Batasan Masalah

Penelitian ini membatasi masalah sebagai berikut :

1. *Clustering* diimplementasikan pada kasus pengelompokkan metode pelatihan sepak bola.
2. Algoritma yang dilakukan pengujian adalah Enhanced K-Means dan Dynamic Cluster.

3. Pengujian dilakukan dengan menggunakan data model dan data kasus pengelompokkan metode pelatihan sepak bola.

1.7 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi sisi keilmuan, menghasilkan algoritma yang mampu meningkatkan hasil *clustering* dengan cara menggabungkan algoritma Enhanced K-Means dan Dynamic Cluster yang memiliki kemampuan untuk menentukan dugaan jumlah *cluster* dan titik pusat awal terbaik.
2. Bagi bidang penelitian, menghasilkan informasi mengenai karakteristik algoritma terhadap karakteristik data tertentu sehingga dapat mempermudah untuk menerapkan algoritma pada kasus nyata yang sesuai dengan karakteristik algoritma.

1.8 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tesis ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menguraikan tentang latar belakang masalah, identifikasi masalah, premis, hipotesis, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Bab ini menguraikan tentang dasar-dasar teori mengenai *clustering* dalam *data mining*, teknik-teknik dalam *data mining* serta evaluasi hasil *clustering*.

BAB 3 METODE PENELITIAN

Bab ini menguraikan tentang penggabungan algoritma Enhanced K-Means dan Dynamic Cluster, data yang digunakan dalam penelitian serta alur atau teknik pengujian dan analisis data.

BAB 4 HASIL PENELITIAN

Bab ini menguraikan tentang hasil penelitian yang disajikan dalam bentuk grafik-grafik dan analisis hasil penelitian.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menguraikan tentang kesimpulan dari penelitian yang telah dilakukan serta saran-saran untuk penelitian selanjutnya.