

BAB 2

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian-penelitian terdahulu yang berhubungan dengan *clustering* yaitu penelitian yang dilakukan oleh Rismawan (2008). Pada penelitian ini, dibangun suatu sistem untuk mengelompokkan data yang ada berdasarkan status gizi dan ukuran rangkanya dengan memasukkan parameter kondisi fisik dari orang tersebut. Pengelompokkan data dilakukan dengan menggunakan metode *clustering* K-Means yaitu dengan mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas. Dari hasil penelitian, dapat disimpulkan bahwa algoritma klasifikasi K-Means dapat digunakan untuk mengelompokkan mahasiswa berdasarkan status gizi dan ukuran kerangka. Dari data yang dilatih, diperoleh 3 kelompok berdasarkan BMI dan ukuran kerangka.

Penelitian yang dilakukan Oyelade (2010) berjudul "*Application of k-Means Clustering algorithm for prediction of Student's Academic Performance*". Penelitian ini menggambarkan algoritma *clustering* k-means sebagai alat yang sederhana dan efisien dalam memonitor perkembangan mahasiswa-mahasiswa. Dengan bantuan metode *data mining*, seperti algoritma *clustering*, memungkinkan untuk menemukan karakteristik kunci dari performansi mahasiswa-mahasiswa dan menggunakan data tersebut untuk memprediksi perkembangannya. Dalam penelitian ini, diimplementasikan algoritma *clustering* k-means tradisional dan pengukuran *euclidean distance* yang

telah dipilih dalam menganalisis skor mahasiswa-mahasiswa.

Penelitian yang berjudul "*Application of AHP and K-Means Clustering for Ranking and Classifying Customer Trust in M-commerce*" yang dilakukan oleh Ibrahim (2011) bertujuan untuk mengeksplor dan mengevaluasi faktor-faktor yang mempengaruhi kepercayaan pelanggan pada website M-commerce (*mobile commerce*) berdasarkan pandangan pelanggan, ketika pelanggan berinteraksi terhadap vendor melalui *wireless network* pada website-website M-commerce. Ada 2 (dua) jenis kuesioner yang dibuat. Yang pertama dibuat untuk pendapat para ahli dan yang kedua didesain untuk mendapatkan data pada tiga (3) website M-commerce. Peringkat akhir faktor-faktor yang didapat, diolah menggunakan AHP dan selanjutnya diolah menggunakan algoritma *clustering k-means*. Hasil dari algoritma *clustering k-means* didapatkan 5 (lima) level kepercayaan pelanggan.

Penelitian yang dilakukan Tajunisha (2010) yang berjudul "*Performance analysis of k-means with different initialization methods for high dimensional data*" bertujuan untuk mengembangkan metode *clustering* yang efektif untuk dataset berdimensi tinggi. Dari semua metode algoritma *clustering*, algoritma k-means merupakan salah satu yang paling dapat dimengerti dalam mengelompokkan dataset kedalam grup-grup. Penelitian ini menggunakan 2 (dua) metode pembandingan, yang pertama adalah metode yang biasa digunakan dan metode yang kedua adalah dengan *Principal Component Analysis (PCA)* untuk reduksi dimensi dan untuk menemukan centroid inti untuk algoritma k-means. Selanjutnya, peneliti

menggunakan pendekatan heuristik untuk mengurangi jumlah kalkulasi jarak dan menetapkan *datapoint* ke dalam *cluster*. Hasil akhir yang didapatkan, terlihat bahwa metode baru yang diusulkan lebih baik daripada metode yang sebelumnya.

Pada penelitian yang dilakukan oleh Wijaya (2010) ini diangkat suatu kasus yaitu menganalisis algoritma K-Means untuk sistem pendukung keputusan penjurusan siswa di MAN Binong Subang. Algoritma K-Means dapat mengelompokkan (segmentasi) data yang mempunyai atribut dan mempunyai jumlah data yang banyak, sehingga dapat dimanfaatkan dalam sistem penentuan penjurusan siswa yang sesuai dengan kemampuan akademik siswa. Adapun tujuan yang ingin dicapai adalah untuk menguji tepat atau tidaknya algoritma *k-means* dalam sistem pendukung keputusan penjurusan siswa. Berdasarkan hasil analisis terhadap algoritma K-Means untuk sistem pendukung keputusan penjurusan, maka kesimpulan yang dapat diambil adalah algoritma K-Means kurang tepat untuk sistem pendukung keputusan penjurusan tetapi algoritma K-Means lebih tepat untuk mengelompokkan data siswa berdasarkan data nilai yang bisa memberikan gambaran untuk penjurusan siswa.

2.2. Penelitian Sekarang

Penelitian ini bertujuan untuk mendapatkan grup-grup berisi jalan-jalan yang memiliki kemiripan. Penelitian ini dilakukan dengan mengambil sampel jalan-jalan besar di Yogyakarta. Penelitian ini dilakukan dengan mengambil sampel jalan-jalan besar di Yogyakarta. Variabel-variabel yang akan digunakan

adalah berupa jenis-jenis usaha yang dijumpai peneliti di jalan-jalan tersebut. Hasil pengumpulan data dimasukkan kedalam bentuk matriks. Langkah-langkah analisis *cluster* ditampilkan secara manual dan hasil didapatkan dengan mengolah data menggunakan *software* MINITAB 14. Dalam menentukan nilai *k* (jumlah *cluster*) yang tepat, pertama-tama, dilakukan analisis *cluster* pada nilai $k=20$, $k=15$, $k=10$, dan $k=5$, dengan membandingkan hasil analisis tersebut, akan didapatkan rentang nilai *k* yang akan dianalisis menggunakan perhitungan *Sum Squared of Errors* (SSE) untuk mendapatkan nilai *k* yang terbaik.

Tabel 2.1. Perbandingan Penelitian Terdahulu dan Sekarang

No	Peneliti dan Tahun	Metode yang digunakan	Tujuan Penelitian
1	Tedy Risnawan (2008)	Algoritma <i>clustering</i> k-means	Membangun suatu sistem untuk mengelompokkan data yang ada berdasarkan status gizi dan ukuran rangkanya dengan memasukkan parameter kondisi fisik
2	Oyelade (2010)	Algoritma <i>clustering</i> k-means	menggambarkan algoritma <i>clustering</i> k-means sebagai <i>tool</i> yang sederhana dan efisien dalam memonitor perkembangan mahasiswa-mahasiswa
3	Othman Ibrahim (2011)	Algoritma <i>clustering</i> k-means dan AHP	Mengeksplor dan mengevaluasi faktor-faktor yang mempengaruhi kepercayaan pelanggan pada website M-commerce (mobile commerce) berdasarkan pandangan pelanggan, ketika pelanggan berinteraksi terhadap vendor melalui wireless network pada website-website M-commerce

Lanjutan Tabel 2.1

No	Peneliti dan Tahun	Metode yang digunakan	Tujuan Penelitian
4	Tajunisha (2010)	<i>Principal Component Analysis</i> (PCA) dan algoritma <i>k-means</i>	Membandingkan dua metode, metode yang biasa digunakan dan metode yang telah dikembangkan
5	Arim Wijaya (2010)	Algoritma <i>clustering k-means</i>	Menguji tepat atau tidaknya algoritma <i>k-means</i> dalam sistem pendukung keputusan penjurusan siswa.
6	Penelitian sekarang	Algoritma <i>clustering k-means</i> dan metode <i>hierarchical clustering</i> .	Menghasilkan pemetaan usaha dan <i>clustering</i> jalan-jalan utama di kota Yogyakarta berdasarkan karakteristik usaha-usaha yang ada.