

BAB III METODE PENELITIAN

3.1 Bahan Penelitian

Data yang digunakan dalam penelitian pengelompokan Sekolah Menengah Kejuruan (SMK) Se-Kabupaten Karawang bersumber dari *website* resmi Pusat Penilaian Pendidikan <https://puspendik.kemdikbud.go.id/hasil-un/> yang dirujuk oleh KEMENDIKBUD dengan alamat <https://puspendik.kemdikbud.go.id>. Data yang diunggah oleh KEMENDIKBUD merupakan data hasil nilai Ujian Nasional SMP, SMA, dan SMK se-Indonesia tahun pelajaran 2018/2019, yang didalamnya memiliki 11 atribut yaitu no, kode, nama satuan pendidikan, NPSN, status, jumlah peserta, bahasa Indonesia, bahasa Inggris, matematika, kompetensi dan rerata nilai. Sehingga data yang digunakan membutuhkan beberapa prosedur sebelum dijadikan bahan penelitian.

3.2 Peralatan Penelitian

Penelitian ini didukung perangkat hardware dan software diantaranya :

Tabel 3.1 Peralatan Penelitian

Perangkat Penelitian	Komponen
Perangkat Hardware	<ol style="list-style-type: none">1. Laptop HP processor AMD Ryzen 5 2500U2. Memory penyimpanan RAM 4GB, dan HDD 1TB dan SSD 128GB
Perangkat Software	<ol style="list-style-type: none">1. Sistem Operasi Windows 102. Software <i>Rapidminer</i>3. Bahasa pemrograman python 3.6.54. <i>Web Open Source Jupyter Notebook</i>

3.3 Lokasi dan Waktu Penelitian

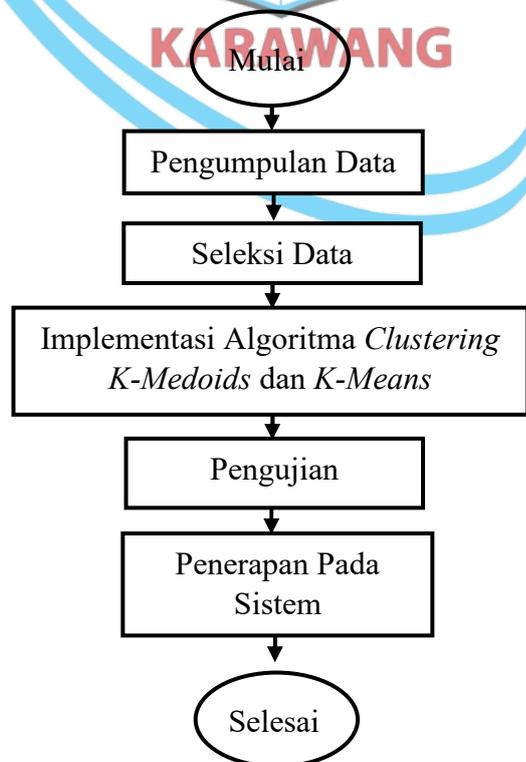
Pelaksanaan pencarian data dilakukan melalui *website* resmi Pusat Penelitian Pendidikan yang diakses pada tanggal 10 Desember 2019. Waktu penelitian dimulai bulan Desember 2019 dilanjut dengan penyusunan proposal Tugas Akhir serta konsultasi dengan dosen pembimbing. Adapun waktu penelitian dalam penyusunan proposal tugas akhir ini sebagai berikut:

Tabel 3.2 Waktu Penelitian

No	Kegiatan	Bulan						
		1	2	3	4	5	6	7
1.	Identifikasi Masalah	■						
2.	Pengumpulan data	■	■					
3.	Seleksi data	■	■	■				
4.	Implementasi Algoritma <i>K-Medoids</i>		■	■	■			
5.	Implementasi Algoritma <i>K-Means</i>			■	■	■		
5.	Pengujian				■	■	■	
6.	Penerapan Pada Sistem						■	■

3.4 Prosedur Penelitian

Penelitian yang akan dilakukan, menggunakan data Ujian Nasional SMK tahun pelajaran 2018/2019. Data tersebut akan diolah menggunakan Algoritma *K-Means* dan *K-Medoids*, lalu dilakukan pengujian terhadap hasil kedua algoritma menggunakan software *Rapidminer*, dan perancangan pada sistem aplikasinya. Berikut adalah gambaran prosedur penelitian yang dilakukan:



Gambar 3.1 Prosedur Penelitian

3.4.1 Pengumpulan Data

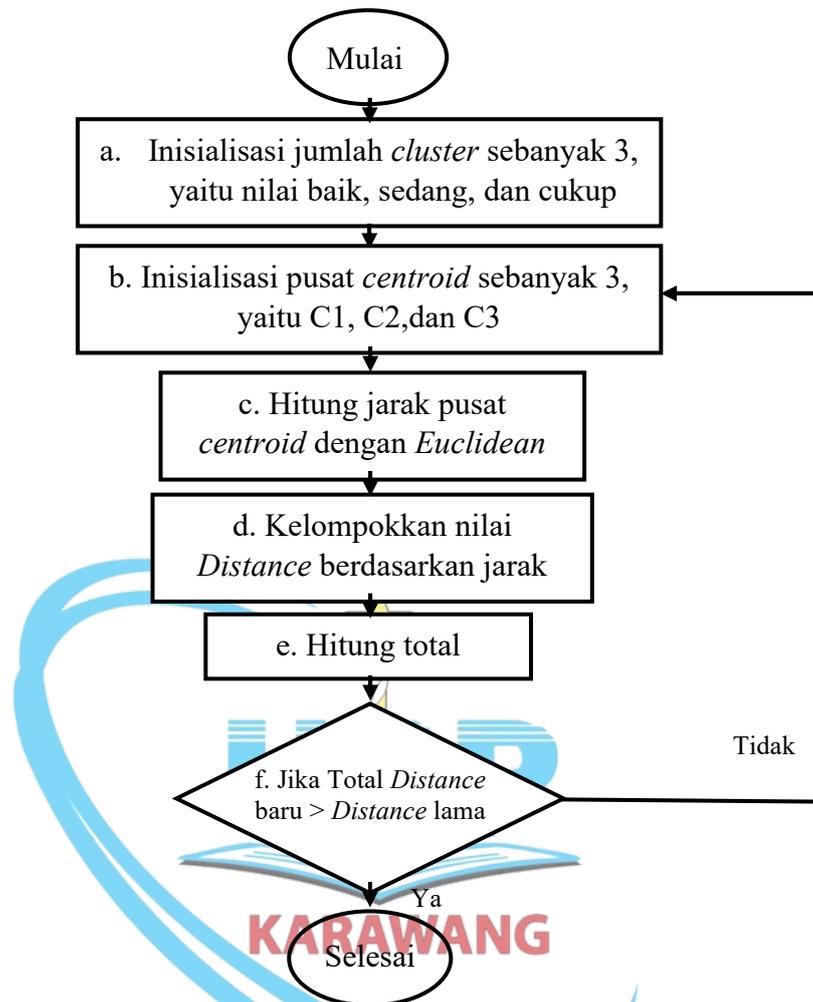
Pengumpulan data dilakukan pada *website* resmi Pusat Penelitian Pendidikan <https://puspendik.kemdikbud.go.id/hasil-un/>, data yang digunakan dalam penelitian merupakan data nilai Ujian Nasional SMK tahun pelajaran 2018/2019. Data ini digunakan karena merupakan data nilai Ujian Nasional kelulusan terbaru pada saat penelitian ini dilakukan. Data yang ditampilkan pada halaman *website* merupakan data nilai SMK se-Indonesia, dimana setiap data tersebut memiliki 11 atribut. . Data tersedia dalam bentuk format file *Microsoft Excel* (.xls). Data lalu dianalisa untuk mengambil sampel yang dijadikan sebagai uji coba perhitungan pada penelitian.

3.4.2 Seleksi Data

Data yang telah diseleksi untuk diambil sampel sebagai uji coba penelitian merupakan data nilai Ujian Nasional SMK di Kabupaten Karawang yang memiliki 11 atribut. Selanjutnya, dilakukan *cleaning* data terhadap atribut yang bertujuan untuk mengurangi *noise* yang dapat mempengaruhi perhitungan, maka dipilih beberapa atribut yang dapat dijadikan acuan dalam proses pengelompokan.

3.4.3 Implementasi Algoritma *Clustering K-Medoids*

Tahap selanjutnya melakukan pengelompokan dengan menggunakan algoritma *clustering* dengan teknik *K-Medoids*. Pengelompokan dilakukan untuk menentukan 3 kategori *cluster* yaitu baik, sedang dan cukup terhadap data sampel uji coba penelitian. Berikut tahapan *flowchart* algoritma *clustering K-Medoids* terhadap dataset nilai Ujian Nasional yang akan dijelaskan pada Gambar 3.2 :



Gambar 3.2 Flowchart Algoritma *K-Medoids*

Pada Gambar 3.2 tahapan *flowchart* implementasi pengelompokan algoritma *K-medoids*, berikut penjelasan dari langkah-langkah tersebut :

- Akan mencari 3 *cluster* yaitu nilai baik, sedang, dan cukup dari dataset nilai Ujian Nasional SMK di Kab. Karawang.
- Mencari titik *centroid* yang akan dijadikan *medoid* awal dipilih secara berurutan atau acak dari dataset, titik *centroid* yang dipilih sejumlah dengan *cluster* yang sudah ditentukan.
- Mengukur jarak dilakukan dengan menggunakan rumus *Euclidean Distance*, yang mengacu pada persamaan 2.1 :

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

Keterangan :

$D(x_2, x_1)$: Jarak variabel x_2 terhadap x_1

j : variabel data ($j = 1, 2, 3, \dots, n$)

p : dimensi data

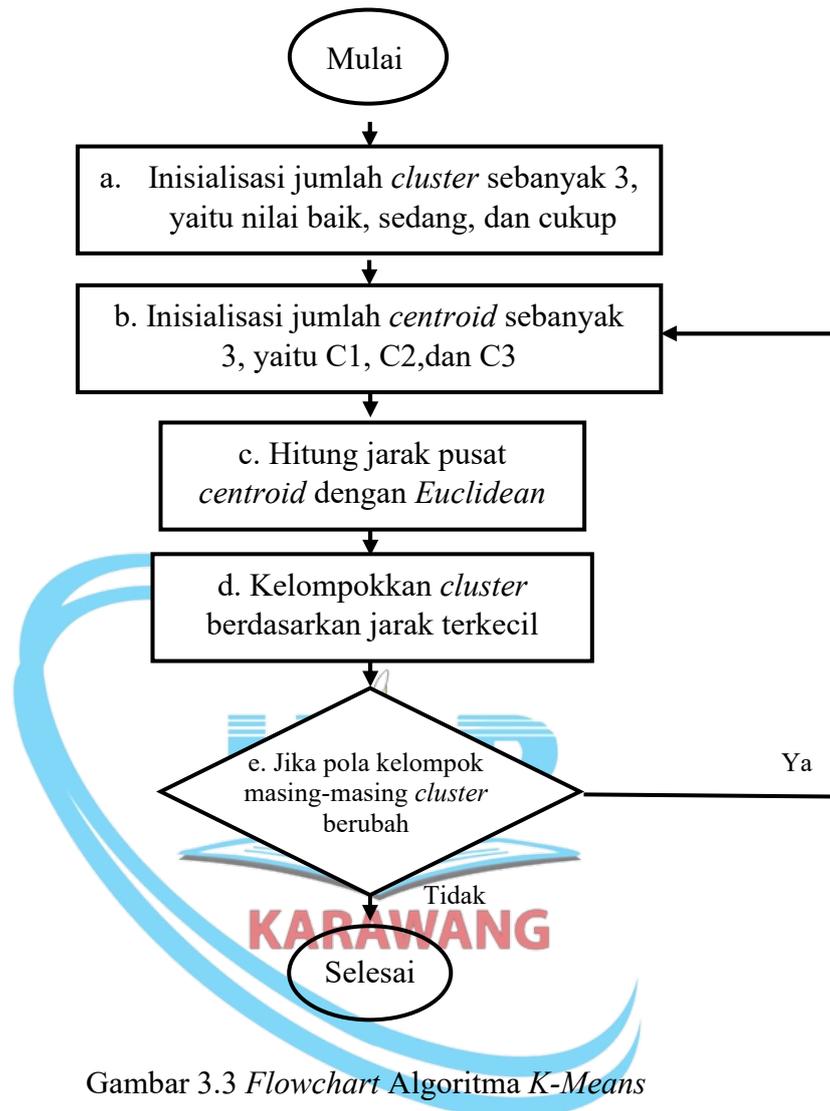
x_2 : variabel data yang diujikan (*testing*)

x_1 : data sampel (*training*)

- d. Kelompokkan jarak berdasarkan jarak terkecil dari setiap data yang telah dihitung, Iterasi *medoid* dilakukan dengan menghitung jarak *Euclidean Distance*, terhadap seluruh data yang dijadikan data proses.
- e. Hitung total perhitungan jarak iterasi pertama, terhadap seluruh data yang dijadikan data proses. Perhitungan jarak dilakukan kembali untuk mendapatkan hasil iterasi ke-2 sebagai perbandingan total simpangan diantara kedua iterasi. Titik *centroid* yang digunakan untuk *medoid* ke-2 menggunakan titik centroid yang baru, yang dipilih secara berurutan atau acak dari dataset.
- f. Perhitungan total simpangan (S) dihitung dari hasil perhitungan jarak iterasi pertama dan kedua. Hasil perhitungan total simpangan, jika total jarak baru lebih besar daripada jarak lama maka iterasi diberhentikan dengan anggota *cluster* diambil dari iterasi terakhir yang terbentuk dari perhitungan jarak. Tetapi apabila hasil total jarak baru lebih kecil dari jarak lama, maka iterasi dilakukan lagi dengan kembali pada point b, sampai hasil total jarak baru lebih besar daripada jarak lama.

3.4.4 Implementasi Algoritma *Clustering K-Means*

Selanjutnya, melakukan tahap implementasi algoritma *clustering K-Means*. Implementasi dilakukan untuk menentukan 3 kategori *cluster* yaitu baik, sedang dan cukup terhadap dataset hasil nilai Ujian Nasional. Berikut langkah-langkah *flowchart* algoritma *clustering K-Means*:



Gambar 3.3 Flowchart Algoritma K-Means

Tahapan algoritma *K-Means* hampir sama dengan tahapan algoritma *K-Medoids* pada point a sampai ke point c. Lanjutan tahapan algoritma *K-Means*:

- a. Iterasi *K-Means* dilakukan dengan mengamati pola kelompok berdasarkan jarak terkecil pada masing-masing *cluster* yang terbentuk.
- b. Perhitungan jarak dilakukan kembali untuk mendapatkan hasil iterasi ke-2 sebagai perbandingan pola kelompok pada *cluster* yang terbentuk. Titik *centroid* yang digunakan untuk *mean* ke-2 menggunakan nilai rata-rata titik *centroid* jarak terkecil yang terbentuk dari iterasi pertama.
- c. Selanjutnya dilakukan perbandingan pola yang terbentuk pada tiap *cluster*. Apabila terjadi perubahan pada pola, maka iterasi dilakukan kembali sampai tidak terjadi perubahan bentuk pola pada masing-masing *cluster*.

- d. Anggota *cluster* diambil dari iterasi terakhir yang terbentuk dari pola perhitungan berdasarkan jarak terkecil.

3.4.5 Pengujian

Pengujian dilakukan menggunakan hasil implementasi dari algoritma *K-Means* dan *K-Medoids* menggunakan *tools Rapidminer*. Pengujian ini bertujuan untuk mengetahui jumlah *cluster* yang optimal serta mencocokkan kesesuaian hasil dari perhitungan implementasi yang dilakukan.

3.4.6 Penerapan Algoritma Pada Sistem

Penerapan algoritma pada sistem menggunakan bahasa pemrograman *python* dengan *web open source Jupyter Notebook*. Data yang tersedia dalam bentuk format file *Microsoft Excel (.xls)*, akan dirubah terlebih dahulu ke dalam bentuk format *Comma Separated Value (.csv)*. Sistem pengelompokan sekolah yang dibuat menerapkan salah satu algoritma yaitu algoritma *k-means*. *Output* yang dihasilkan pada tabel *clustering* sekolah akan ditampilkan dalam bentuk *view* tabel *clustering*.

