

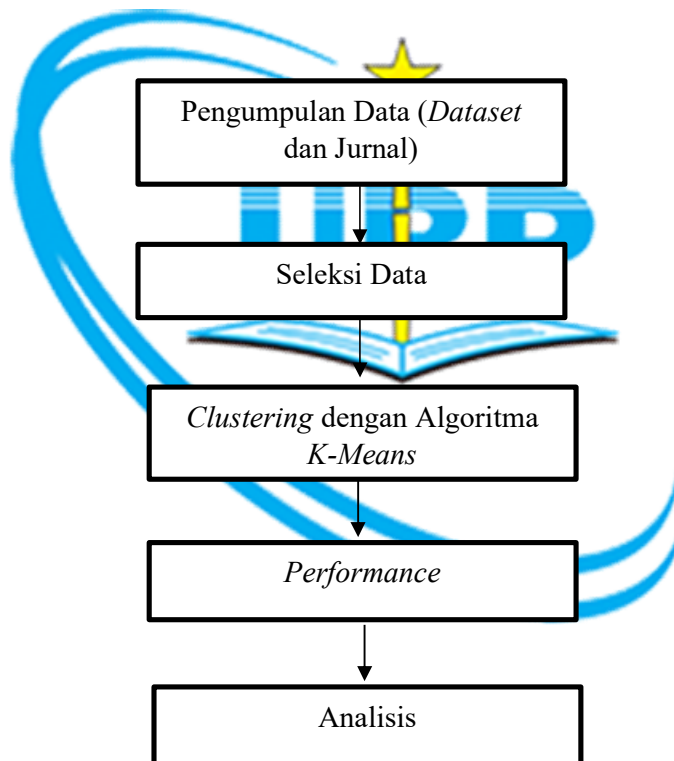
BAB III

METODE PENELITIAN

3.1 Gambaran Umum Penelitian

Gambaran umum dari penelitian ini yaitu perancangan yang akan dibuat oleh peneliti. Dalam penelitian ini juga dipergunakan alat uji penelitian berupa sebuah aplikasi pengujian data mining dengan Rapidminer.

Penelitian ini akan dilaksanakan sesuai rancangan penelitian seperti yang ditunjukkan pada gambar 3.1



Gambar 3. 1 Rancangan Penelitian

3.2 Pengumpulan Data

Pada penelitian ini metode pengumpulan data yang digunakan adalah studi dokumen. Studi dokumen merupakan Pengambilan data melalui dokumen tertulis maupun elektronik dari lembaga/institusi.

Data dalam penelitian ini berasal dari sumber data utama dan sumber data sekunder

a. Sumber utama

Sumber data utama ini diperoleh dari situs website Badan Pusat Statistika yaitu <https://www.bps.go.id>. Data tersebut yaitu tabel persentase penduduk buta aksara (buta huruf) dengan kriteria data meliputi nama provinsi, tahun, umur/umur diantaranya 15+, 15-45, 45+. Tabel bisa dilihat pada lampiran 1.

b. Sumber sekunder

Sumber sekunder ini diperoleh dari studi literatur yaitu dengan mencari materi yang berhubungan dengan permasalahan, perancangan, metode *K-Means Clustering*, penunjang keputusan dan buta aksara guna mempermudah proses implementasi. Studi literatur diperoleh melalui buku, jurnal diinternet dan pada penelitian sebelumnya.

3.3 Seleksi Data

Seleksi Data merupakan proses pemilihan data yang akan digunakan dalam proses data mining. Dalam penelitian ini penulis hanya menggunakan data tahun 2017 saja dan untuk data ditahun sebelumnya tidak dipakai. Data asli yang didapat oleh penulis terdiri dari tahun 2011 sampai tahun 2017 dengan 5 buah atribut yaitu atribut provinsi, umur 15+, umur 15-45, umur 45+ dan atribut tahun. Setelah dilakukan seleksi data maka hanya 4 buah atribut yang dipakai yaitu atribut provinsi, Umur 15+, Umur 15-45, dan Umur 45+, Sehingga didapatkan dataset di tahun 2017 terlihat pada Tabel 3.1 dibawah ini

Tabel 3. 1 Dataset persentase penduduk buta huruf

Provinsi	2017		
	Persentase Penduduk Buta Huruf (Persen)		
	15+	15-44	45+
Aceh	2.06	0.13	6.82
Sumatera Utara	1.11	0.38	2.64
Sumatera Barat	1.15	0.19	2.87
Riau	0.83	0.15	2.62
Jambi	1.91	0.29	5.49
Sumatera Selatan	1.46	0.25	4.04
Bengkulu	2.1	0.27	6.19
Lampung	3.11	0.16	8.81
Kep. Bangka Belitung	2.21	0.7	5.53

Tabel 3. 2 Dataset persentase penduduk buta huruf (lanjutan)

Provinsi	Tahun 2017 Persentase Penduduk Buta Huruf (Persen)		
	15+	15-45	45+
Kep. Riau	1.17	0.29	3.88
Dki Jakarta	0.33	0.06	0.92
Jawa Barat	1.77	0.26	4.71
Jawa Tengah	6.61	0.36	15.38
Di Yogyakarta	5.36	0.2	12.38
Jawa Timur	8.18	1.01	18.27
Banten	2.43	0.21	7.97
Bali	7.1	0.4	17.58
Nusa Tenggara Barat	12.86	3.2	32.92
Nusa Tenggara Timur	8.32	3.08	18.92
Kalimantan Barat	7.52	1.76	20.2
Kalimantan Tengah	0.92	0.23	2.64
Kalimantan Selatan	1.6	0.15	4.65
Kalimantan Timur	1.04	0.15	3.09
Kalimantan Utara	4.86	1.38	13.55
Sulawesi Utara	0.24	0.19	0.32
Sulawesi Tengah	2.31	0.84	5.39
Sulawesi Selatan	8.35	2.03	20.28
Sulawesi Tenggara	5.68	1.14	16.41
Gorontalo	1.56	0.5	3.81
Sulawesi Barat	7.21	3.16	16.77
Maluku	0.87	0.77	1.09
Maluku Utara	1.32	0.41	3.62
Papua Barat	2.84	1.94	5.55
Papua	26.11	24.66	30.46
Indonesia	4.5	0.94	11.08

Sumber: BPS 2017

3.4 *Clustering (dengan Algoritma K-Means)*

Setelah melakukan proses *data cleaning*, tahap selanjutnya adalah pengolahan *dataset* dengan menggunakan algoritma *K-Means*. Langkah-langkahnya sebagai berikut:

1. Menentukan k buah *cluster*.
2. Pilih sejumlah k buah objek dengan cara acak yang nantinya akan dijadikan titik *centroid cluster*.
3. Menentukan k buah *centroid* (titik tengah).
4. Kelompokkan objek ke *centroid cluster* yang paling dekat sesuai jarak objek ke *centroid* (*Euclidean distance*).
5. Menghitung kembali semua nilai titik *centroid*.
6. Mengulang langkah 3,4 dan 5 sampai titik *centroid* nilainya tidak berubah lagi

3.5 *Performance*

Selanjutnya yaitu tahapan *performance* atau kinerja Algoritma *K-Means*. Data yang akan diuji berupa *dataset* menggunakan *tools* Rapidminer dengan metode algoritma *K-Means*. Tahapan ini nantinya akan menghasilkan 3 *cluster* yaitu: tingkat buta aksara tinggi, sedang dan rendah.

3.6 *Analisa*

Selanjutnya, *cluster* tingkat buta aksara akan dianalisa, apakah hasil *cluster* dengan pengujian *tools* Rapidminer dan pemrograman PHP memiliki hasil yang sama dengan perhitungan manual atau tidak. Dari ketiga pengujian tersebut, maka dapat disimpulkan algoritma *K-Means* dapat diterapkan untuk pengelompokan data persentase buta aksara.