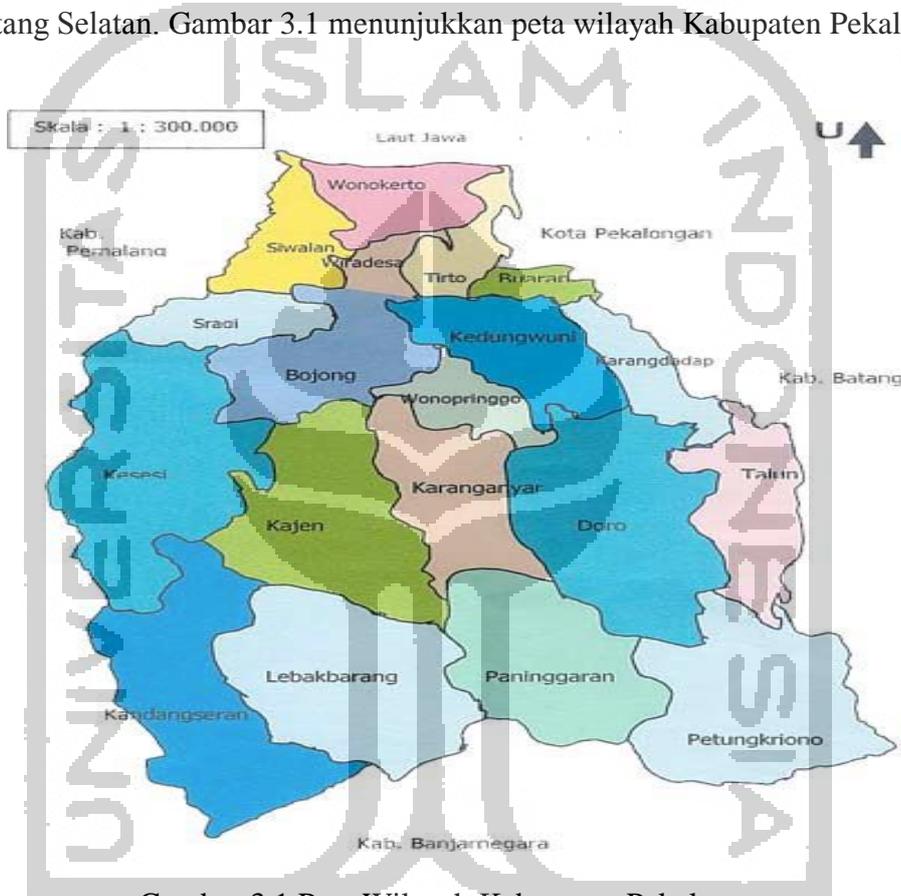


BAB III

Metodologi Penelitian

3.1 Lokasi Penelitian

Lokasi Penelitian dilakukan di Kabupaten Pekalongan yang merupakan salah satu bagian dari wilayah Provinsi Jawa Tengah, terletak diantara 109°-109° 78" Bujur Timur dan 6°-7°23" Lintang Selatan. Gambar 3.1 menunjukkan peta wilayah Kabupaten Pekalongan.



Gambar 3.1 Peta Wilayah Kabupaten Pekalongan

Berdasarkan Gambar 3.1, Kabupaten Pekalongan berbatasan dengan Kota Pekalongan dan Laut Jawa disebelah Utara, Kabupaten Banjarnegara disebelah Selatan, Kota Pekalongan dan Kabupaten Batang disebelah Timur, serta Kabupaten Pemalang disebelah Barat. Kondisi wilayah Kabupaten Pekalongan terdiri dari tanah sawah 29,37%, tanah tegalan/kebun 11,70%, hutan rakyat 4,40%, tanah rumah, bangunan dan halaman 14,38%, perkebunan 2,92%, tambak 1,00%, rawa 0,02% dan tadah hujan 4.57 %. Wilayah dataran rendah (0-50 m dpl), meliputi Kecamatan Wonokerto, Sragi, Siwalan, Kedungwuni, Tirta, Karangdadap, Wiradesa, Wonopringgo, dan Buaran. Wilayah dataran sedang (50-400 m dpl), meliputi Kecamatan Bojong, Karanganyar, Kesesi, Doru, Kajen,

dan Talun. Sedangkan wilayah dataran tinggi (> 400 m dpl), meliputi Kecamatan Kandangserang, Lebakbarang, Paninggaran, dan Petungkriyono. Tabel 3.1 menunjukkan luas wilayah per-kecamatan di Kabupaten Pekalongan.

Tabel 3.1 Luas Wilayah PerKecamatan di Kabupaten Pekalongan

No	Kecamatan	Luas (Km ²)	Desa/Kelurahan
1	Kandangserang	60,55	14
2	Paninggaran	92,99	15
3	Lebakbarang	58,20	11
4	Petungkriyono	73,58	9
5	Talun	58,57	10
6	Doro	68,45	14
7	Karanganyar	63,48	15
8	Kajen	75,15	25
9	Kesesi	68,52	23
10	Sragi	32,40	17
11	Siwalan	25,91	13
12	Bojong	40,06	22
13	Wonopringgo	18,80	14
14	Kedungwuni	22,94	19
15	Karangdadap	20,99	11
16	Buaran	9,54	10
17	Tirto	17,39	16
18	Wiradesa	12,71	16
19	Wonokerto	15,90	11
Jumlah		836,13	285

Berdasarkan Tabel 3.1, Kabupaten Pekalongan terbagi menjadi 19 wilayah kecamatan terdiri dari 285 desa/kelurahan, dengan luas wilayah keseluruhan adalah ± 836,13 km². Secara topografi Kabupaten Pekalongan terletak pada kawasan kaki Pegunungan Dieng yang terdiri atas wilayah pantai, wilayah dataran rendah, dan wilayah pegunungan dengan ketinggian 0 meter sampai dengan 1.294 meter di atas permukaan laut.

Wilayah pantai terletak di bagian utara, wilayah dataran rendah terletak di bagian barat, tengah dan timur, wilayah pegunungan berada di bagian selatan.

3.2 Jenis Data

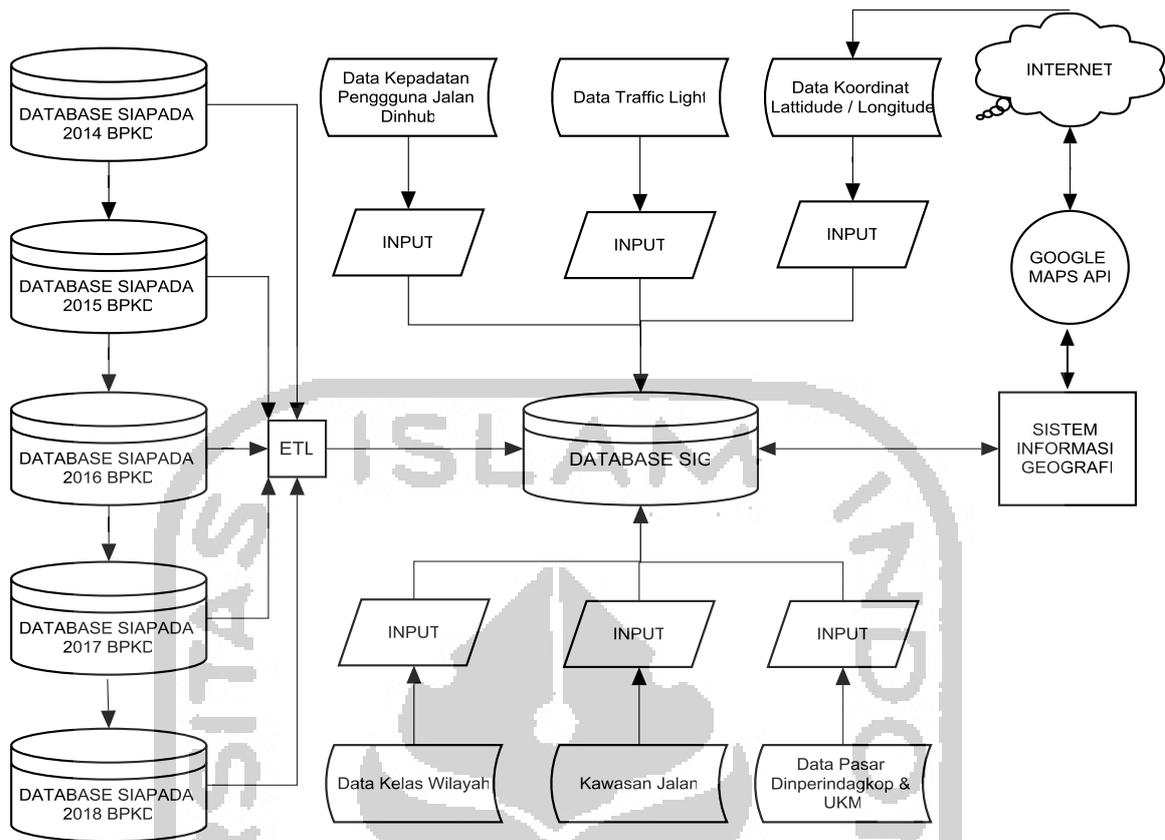
Terdapat dua jenis data yang digunakan dalam penelitian ini: data spasial dan data non spasial. Kedua jenis data ini akan dipaparkan sebagai berikut.

Data spasial adalah gambaran nyata suatu wilayah yang terdapat di permukaan bumi. Data spasial pada umumnya direpresentasikan berupa grafik, peta, gambar dengan format digital dan disimpan dalam bentuk koordinat x,y (vektor) atau dalam bentuk image (raster) yang memiliki nilai tertentu. Data koordinat dan peta didapatkan dari aplikasi SIG yang dibangun menggunakan bahasa pemrograman PHP dan basisdata MySQL. Aplikasi ini terhubung dengan Google Maps API.

Data non spasial adalah data berbentuk tabel yang berisi informasi yang dimiliki oleh obyek data spasial. Data tersebut berbentuk data tabular yang saling terintegrasi dengan data spasial yang ada. Untuk data non spasial seperti data reklame, kawasan jalan, dan kelas wilayah didapat dari studi lapangan. Hal ini dilakukan dengan mengambil informasi dari basisdata SIAPADA (Sistem Informasi Administrasi Pajak Daerah) di Badan Pengelolaan Keuangan Daerah (BPKD) Kabupaten Pekalongan. Sedangkan untuk data pasar diambil dari Dinperindagkop dan UKM. Data tingkat keramaian pengguna jalan diambil dari data survei CTMC (*Classified Turning Movement Counting*) pada Buku Kinerja Keselamatan Transportasi Jalan di Kabupaten Pekalongan (Nugraha, 2016).

3.3 Teknik Pengumpulan Data

Penelitian ini dilakukan dengan mengumpulkan data pajak reklame dari basisdata pembayaran pajak pada BPKD Kabupaten Pekalongan, dari basisdata tersebut dilakukan proses *extract, transform, load* (ETL) menuju basisdata aplikasi SIG yang nantinya dijadikan sebagai bahan analisis. Untuk kebutuhan analisis, ada data-data lain yang perlu dimasukkan ke dalam basisdata, diantaranya adalah data koordinat reklame, data pasar beserta koordinatnya, data *traffic light* beserta koordinatnya, data kawasan jalan, data kelas wilayah dan data volume kendaraan. Gambar 3.2 menunjukkan diagram alir pengumpulan data.



Gambar 3.2 Diagram Alir pengumpulan data

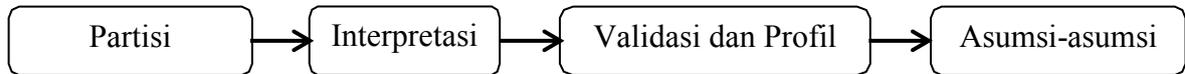
Berdasarkan Gambar 3.2, pengumpulan data penelitian diambil dari database SIAPADA melalui proses ETL (*Extract, Transform, Load*). Data-data lain terkait penelitian seperti data kelas wilayah, kawasan jalan, pasar, *traffic light*, volume kendaraan dan koordinat diinput melalui aplikasi.

3.4 Teknik Penelitian

3.4.1 Analisis Cluster

Para peneliti dan akademisi sering dihadapkan pada situasi yang memungkinkan mereka dapat mendefinisikan objek yang homogen dengan banyak cara. Objek bisa merupakan individu, perusahaan, produk atau bahkan tingkah laku seseorang. Dalam hal ini, interpretasi yang tepat tidaklah mungkin didapat tanpa metodologi yang tepat pula. Oleh karenanya, diperlukan suatu teknik atau cara untuk mengatasinya. Analisis *cluster* merupakan teknik yang biasa dipakai untuk maksud tersebut. Analisis *cluster* meliputi algoritma pengelompokan objek yang berbeda-beda. Tujuan utama analisis ini adalah mengelompokkan objek-objek atau kasus-kasus ke dalam satu atau lebih *cluster* (kelompok) berdasarkan kesamaan karakteristik di antara objek-objek tersebut, sehingga

objek-objek yang berada dalam satu *cluster* akan mempunyai kemiripan satu sama lain. Analisis *cluster* termasuk analisis *interdependensi multivariat* yaitu analisis *multivariat* yang memperhatikan struktur keterkaitan antar variabel independen saja (tidak ada variabel dependen). Gambar 3.3 menunjukkan diagram alir tahapan analisis *cluster*.



Gambar 3.3 Diagram Alir Tahapan Analisis *Cluster*

a. Tahap Partisi

Merupakan teknik dalam pengelompokan yang merupakan dasar analisis. Beberapa hal yang perlu diperhatikan pada tahap ini adalah: Pertama, tentukan objek dan variabel yang akan dianalisis. Tentukan jumlah *cluster* yang akan dibentuk. Kedua, tentukan algoritma pengelompokan objek. Ketiga, tentukan ukuran similaritas yang akan digunakan.

b. Tahap Interpretasi

Pada tahap ini, hasil pengelompokan berupa *cluster-cluster* akan diinterpretasikan sesuai nilai karakteristik yang terkandung dalam objek-objeknya. Biasanya, interpretasi dilakukan berdasarkan nilai rata-rata dari karakteristik objek dalam *cluster*. Hasil interpretasi berupa label atau nama dari masing-masing *cluster*.

c. Tahap Validasi dan Profil

Validasi dilakukan untuk menguji bahwa hasil pengelompokan dapat mewakili populasi penelitian, sifatnya stabil dan berlaku umum untuk objek lainnya. Sedangkan profil, menyangkut penjelasan karakteristik dari setiap *cluster* yang telah diinterpretasikan. Sehingga, perbedaan dimensi antar *cluster* menjadi jelas.

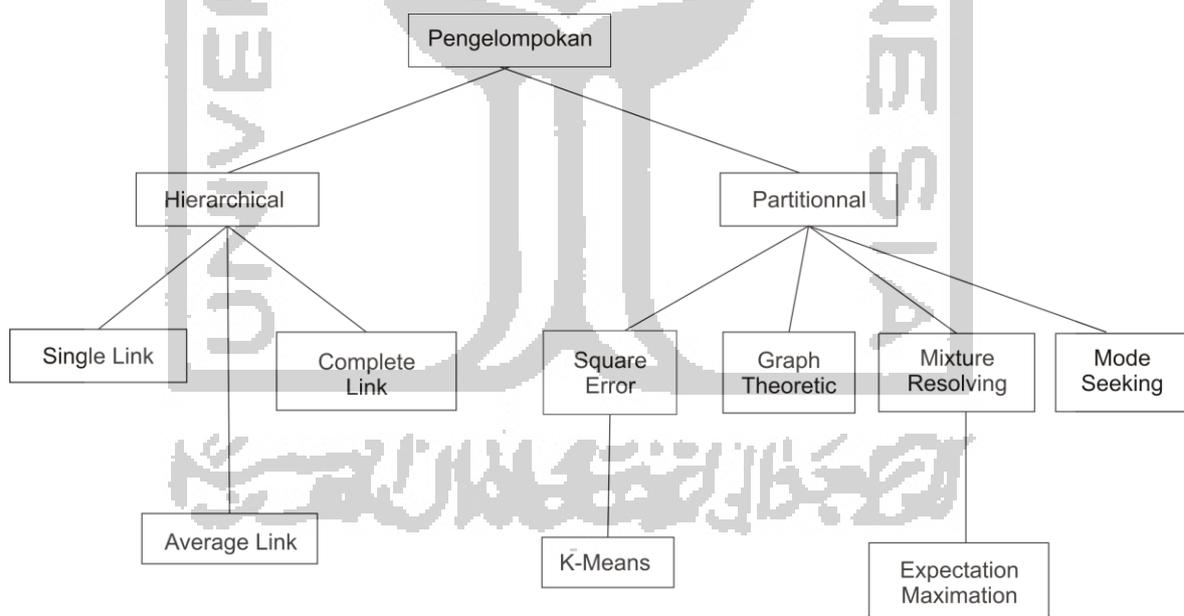
d. Asumsi-Asumsi

Agar terbentuk suatu *cluster* yang baik, yaitu *cluster* yang mempunyai kesamaan yang tinggi antar anggota dalam satu *cluster* (*within cluster*) dan perbedaan yang tinggi antara *cluster* yang satu dengan *cluster* yang lain (*between cluster*), maka haruslah dipenuhi asumsi-asumsi seperti sampel yang diambil benar-benar bisa mewakili populasi yang ada (representatif). Kemudian mempunyai sifat *multikolinieritas*, yakni kemungkinan adanya

korelasi antar objek, sebaiknya tidak ada. Atau seandainya ada, besar multikolinieritas tersebut tidaklah tinggi (misalnya di atas 0,5).

3.4.2 K-means Cluster

Data *Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hierarki) data *clustering* dan *non-hierarchical* (non hierarki) data *clustering*. *K-means* merupakan salah satu metode data *clustering* non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Tujuan dari data *clustering* ini adalah untuk meminimalisasikan fungsi obyektif yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Gambar 3.4 menunjukkan teknik pengelompokan.



Gambar 3.4 Teknik Pengelompokan (Han dkk, 2012)

Berdasarkan Gambar 3.4, pengelompokan data digolongkan menjadi dua yaitu *Hierarchical* dan *Partiionnal*. Pengelompokan *Hierarchical* terdiri dari *Single Link*, *Average Link*, dan *Complete Link*, sedangkan *k-means* merupakan pengelompokan

Partitional. Prinsip dasar *k-means* adalah meminimumkan jumlah kuadrat error (*square error*) dari seluruh kluster yang ada.

3.4.3 Algoritma *K-means*

Algoritma *K-means* merupakan algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar obyek dengan atribut tertentu ke dalam kelompok-kelompok (*cluster*) sebanyak *K*. Pada algoritma *K-means*, jumlah *cluster* *K* sudah ditentukan lebih dahulu.

Setiap *cluster* memiliki titik pusat dan anggota-anggota dari satu *cluster* dipilih berdasarkan jarak dari titik pusat *cluster* terdekat. Penentuan keanggotaan dan titik pusat *cluster* kemudian menjadi tidak mudah, karena penambahan satu anggota pada lokasi yang signifikan akan merubah lokasi titik pusat *cluster*, dan status keanggotaan harus ditinjau kembali, perubahan keanggotaan kemudian akan kembali merubah lokasi titik pusat dan seterusnya keanggotaan mungkin akan berubah lagi. Karena itu, proses penentuan titik pusat dan keanggotaan *cluster* harus dilakukan dalam *iterasi* (perulangan) hingga posisi titik pusat dan anggota-anggota *cluster* benar-benar stabil. Langkah-langkah pada algoritma *K-means* adalah sebagai berikut:

1. Menentukan *K* dimana *K* adalah banyak *cluster* yang ingin dibentuk.
2. Pilih *K* buah data secara acak dari set data yang ada sebagai pusat *cluster* awal.
3. Langkah ke-3 ini lakukan :
 - a. Menentukan jarak setiap data dengan masing-masing pusat *cluster* awal pada langkah 2. Disini akan terlihat data akan menjadi anggota dari *cluster* ke-*k*. Untuk menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus (2.1) korelasi antar dua objek *Euclidean Distance*
 - b. Menghitung *BCV* (*Between Cluster Variation*)

$$BCV = \frac{1}{n_k} \sum_{i=1}^k d(m_i, m_i) \dots\dots\dots (3.1)$$

Dimana *BCV* adalah rata – rata dari *centroid*, *k* adalah jumlah *cluster*, *i* adalah nama yang mewakili cluster yang dibentuk dan *m_i* merupakan jumlah anggota dari cluster ke-*i*

c. Menghitung WCV (*Within Cluster Variation*)

$$WVC = \sum_{i=1}^k \sum a_{ik} d(X_k, m_i)^2 \dots\dots\dots (3.2)$$

Dimana WCV = Jarak antara anggota dalam *Cluster*, sedangkan k= Jumlah data, i = Jumlah *cluster*, a_{ik} = keanggotaan data ke-k ke *cluster* ke-i, m_i = Nilai *centroid* ke-i dan X_k = Data ke-k.

d. Menghitung rasio

$$r = \frac{BCV}{WCV} \dots\dots\dots (3.3)$$

Dimana r = Rasio, WCV = Jarak antara anggota dalam *Cluster* dan BCV = rata-rata dari *centroid*.

4. Perbaharui pusat kelompok

$$m_i = \frac{\sum C_i}{n_i} \dots\dots\dots (3.4)$$

Dimana m_i = *Centroid* ke-i, $\sum C_i$ = Jumlah nilai *cluster* ke-i, n_i = Jumlah data setiap *cluster* ke-i

Jika ada data yang berpindah *cluster*, maka proses dilanjutkan ke iterasi selanjutnya, jika tidak maka data sudah stabil dan proses dihentikan.

3.4.4 Akses Validitas Hasil

Masalah utama dalam validasi analisis *cluster* ialah menentukan banyaknya *cluster* dan penentuan metode *cluster* yang tepat. Ada beberapa cara yang dapat menjadi petunjuk dalam penentuan banyaknya *cluster* diantaranya yaitu: *Pertama*, pertimbangan teoretis atau konseptual bisa diusulkan untuk menentukan berapa banyaknya *cluster* yang sebenarnya. *Kedua*, dengan menggunakan pengclusteran hierarkhi, jarak dimana *cluster* digabung bisa dipergunakan sebagai kriteria. Informasi ini bisa diperoleh dari *agglomeration schedule*. Bila nilai dalam kolom *coefficient*, tiba-tiba menjadi dua kali

antara tahapan sebelumnya. *Ketiga*, suatu pengelompokan (*cluster*) seharusnya berguna/bermanfaat. “Bila suatu *cluster* hanya memiliki 1 anggota maka *cluster* tersebut lebih baik digabung dengan *cluster* lain yang memiliki jarak terdekat” (Supranto, 2004). Untuk menentukan jumlah *cluster* yang optimal, penelitian ini menggunakan library NBClust pada RStudio dengan menggunakan semua indeks yang ada (Charrad, 2014).

