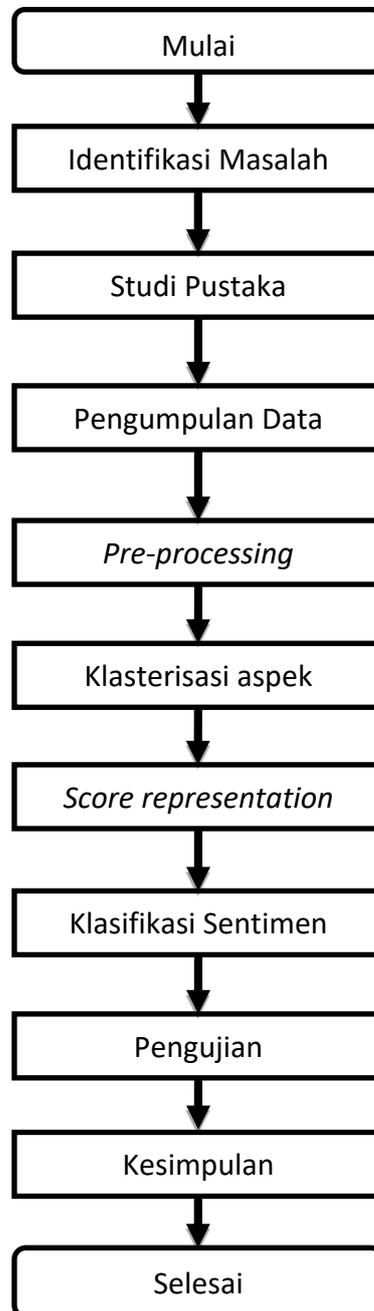


BAB 3 METODOLOGI

Gambaran tahapan-tahapan untuk menyelesaikan penelitian bisa dilihat pada Gambar 3.1.



Gambar 3.1 Alur metodologi penelitian

3.1 Identifikasi Masalah

Dimulai dari mengambil suatu permasalahan yang ada di sekitar kemudian memperjelas permasalahan tersebut hingga ditentukan objek penelitian. Lalu melihat potensi dan kebermanfaatannya penelitian ini.

3.2 Studi Pustaka

Dari permasalahan yang sudah didefinisikan, kemudian dicari sumber-sumber bacaan untuk menunjang penelitian ini. Pada tahap ini dimulai dengan mencari berbagai penelitian sebelumnya tentang topik penelitian yaitu analisis sentimen secara umum dan analisis sentimen tingkat aspek secara khusus, kemudian bagaimana kondisi terkini terkait penelitian tersebut. Setelah didapat satu rujukan utama untuk penelitian, kemudian dicoba untuk mengimplementasikannya dengan sedikit perbedaan pada metode yang digunakan dan objek penelitiannya menyesuaikan dengan lingkungan. Selain pustaka untuk rujukan utama, juga dicari sumber-sumber yang menjelaskan tentang perkembangan metode yang akan digunakan dan hal-hal lain yang akan membantu penyelesaian untuk penelitian ini. Sumber untuk pustaka tersebut di dapat dari artikel ilmiah, buku, situs web, dan jenis penelitian lainnya.

3.3 Pengumpulan Data

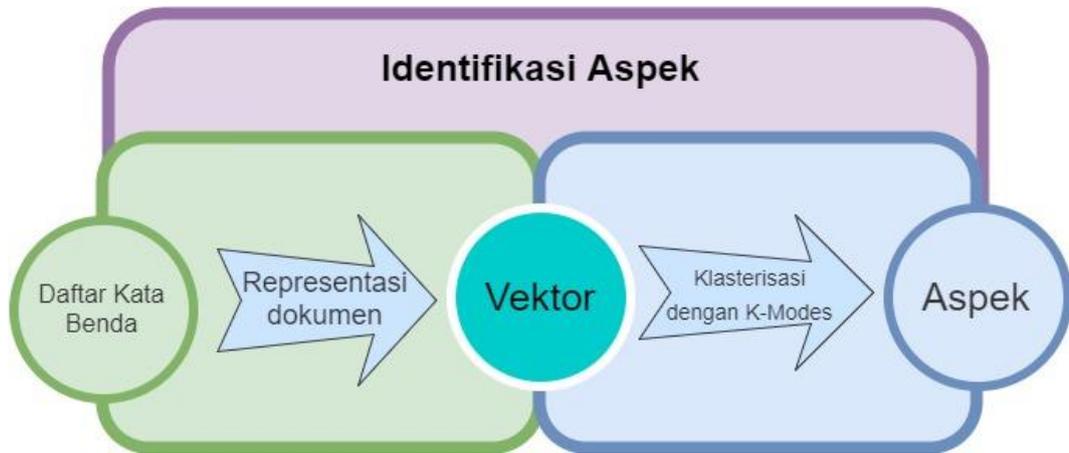
Untuk data yang akan dijadikan data latih dan data uji, diambil dari ulasan yang ada di situs TripAdvisor. Menggunakan salah satu fitur dari TripAdvisor, pertama disaring terlebih dahulu dengan ulasan Bahasa Indonesia. Setelah itu data diambil dengan cara *crawling* menggunakan program Python. Setelah data ulasan terkumpul, dilakukan pemberian label terhadap setiap ulasan apakah termasuk positif atau negatif.

3.4 Pre-processing

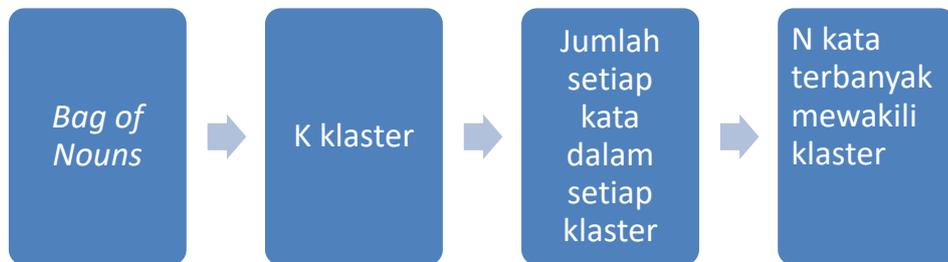
Suatu pemrosesan yang dilakukan untuk memperbaiki data yang akan diolah sebelum bisa digunakan oleh proses klusterisasi atau pengelompokan. Dimulai dengan menyamarkan huruf menjadi huruf kecil, mengambil karakter yang berupa alfabet, kemudian menghapus kata yang sering muncul dan dianggap tidak memiliki makna yang berarti (*stopwords removal*), kemudian menjadikan data tersebut terbagi ke dalam data yang lebih kecil berupa token atau kata. Selanjutnya dengan *POS-Tagging* mengambil kata yang termasuk kata benda (Nomina) dan dimasukkan ke dalam daftar kata benda yang selanjutnya akan digunakan sebagai representasi kalimat sehingga dapat dilakukan klusterisasi atau pengelompokan.

3.5 Identifikasi Aspek

Dengan menggunakan pendekatan untuk membuat daftar kata yang hanya terdiri dari kata benda, akan membantu algoritme klusterisasi untuk mendapatkan hasil yang lebih sesuai untuk mewakili setiap kluster atau kelompok dibanding dengan menggunakan seluruh kata untuk dikluster (Farhadloo dan Rolland, 2013). Algoritme klusterisasi yang digunakan adalah K-Modes. Gambar 3.2 dan Gambar 3.3 menggambarkan proses untuk mengidentifikasi aspek.



Gambar 3.2 Diagram alir dari proses identifikasi aspek



Gambar 3.3 Diagram alir rincian identifikasi aspek

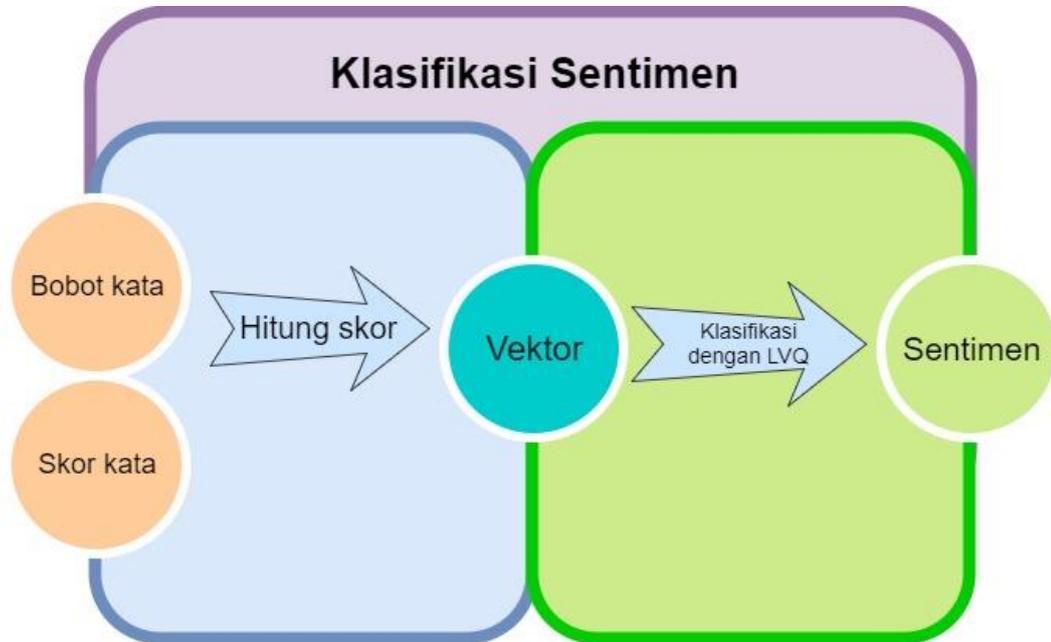
3.6 Klasifikasi Sentimen

Setelah aspek yang dominan diulas didapat, sentimen dari setiap ulasan yang mengandung aspek tersebut diidentifikasi sentimennya. Untuk ekstraksi fitur, digunakan *score representation* seperti yang diusulkan oleh Farhadloo dan Rolland (2013). Namun karena kelas dalam penelitian ini hanya positif dan negatif maka yang diambil untuk *score representation* ini hanya untuk yang positif dan negatif pada Persamaan 2.3 dan Persamaan 2.4. Sehingga, setiap kalimat akan direpresentasikan dengan vektor 2-dimensi ($S = [S^+, S^-]^T$) sebagai ekstraksi fitur.

Satu kalimat terdiri dari dua komponen yaitu nilai positif (S^+) dan nilai negatif (S^-). Nilai tersebut didapat dari penjumlahan hasil perkalian bobot setiap kata (w) dengan *score representation* masing-masing kata. Untuk bobot (w) didapat dengan mengambil jumlah kemunculan kata. Untuk *score representation* setiap nilai positif, dan negatif, didapat dengan membagi frekuensi suatu kata pada kelasnya (positif atau negatif sesuai dengan pelabelan manual) dengan total frekuensi kata tersebut.

Selanjutnya dibuat model jaringan LVQ dengan dua *input*, yaitu $[S^+, S^-]^T$ dan dua *output*, yaitu kelas positif dan kelas negatif. Secara acak dipilih dua vektor

untuk inialisasi bobot awal dan masing-masing mewakili *output* atau kelas yang ada (positif dan negatif). Dalam percobaan nantinya, akan mencoba untuk menggunakan LVQ2. Gambar 3.4 menunjukkan proses untuk mengklasifikasi sentimen.



Gambar 3.4 Diagram alir dari klasifikasi sentimen

3.7 Pengujian

Untuk melihat bagaimana kinerja dari metode yang diterapkan, maka dihitunglah beberapa macam nilai evaluasi. Pertama adalah melihat seberapa tepat suatu dokumen berada pada suatu kluster dengan menggunakan *Silhouette Coefficient*. Kedua adalah melihat kepada hasil analisis sentimen oleh sistem dan hasil analisis sentimen yang telah diberi label atau kelas sebelumnya dengan manual, diantaranya *precision* dan *recall*, kemudian didapat *f1-score*.

3.8 Kesimpulan

Di akhir, dilakukan pengambilan kesimpulan bagaimana penerapan metode ini berjalan dan membandingkan hasilnya dengan penelitian dari Mohsen dan kawan-kawan. Kemudian, diberikan saran untuk penelitian selanjutnya sehingga bisa diperbaiki hal-hal yang bisa meningkatkan hasil penelitian selanjutnya.