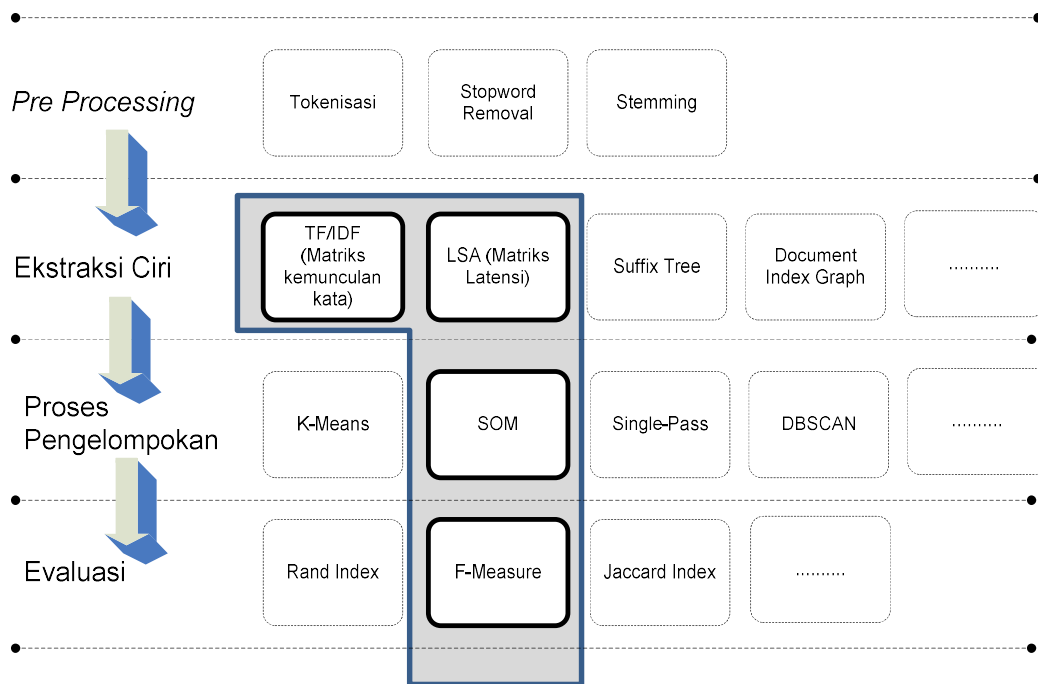


BAB 3 KERANGKA PENELITIAN

3.1 Kerangka Penelitian

Kerangka berpikir yang digunakan pada penelitian ini didasarkan pada tahapan proses kluster secara umum. Berdasarkan tahapan-tahapan tersebut terdapat beberapa hal yang menjadi fokus dan beberapa hal lainnya bukan merupakan fokus dalam penelitian ini :



Keterangan :
 ——— Diteliti
 - - - - - Tidak Diteliti

Gambar 3.1 : Kerangka berpikir.

Sebuah proses pengelompokan dokumen selalu didasarkan pada ciri yang terkandung pada objek dokumen. Selanjutnya, ciri suatu dokumen dapat diidentifikasi dari kemunculan kata maupun latensi. Ciri yang terkandung dalam suatu dokumen harus dapat disajikan dalam model representasi data yang memungkinkan untuk dikluster. Matriks kemunculan kata dan matriks latensi merupakan model representasi data yang bisa diterapkan pada proses kluster.

Ciri kemunculan kata dapat dilihat pada matriks kemunculan kata, sedangkan ciri latensi dapat dilihat pada matriks latensi. Setelah masing-masing ciri diperoleh, proses selanjutnya adalah membentuk Kluster berdasarkan pola masing-masing ciri dengan menggunakan pendekatan SOM. Hasil proses Kluster selanjutnya diukur performanya menggunakan *F-Measure*, yang merupakan kombinasi antara berapa nilai presisi dan *recall* dari masing-masing ciri.

3.2 Data Penelitian

Data penelitian diambil dari arsip konten yang dimiliki oleh beberapa penyedia konten yang relevan selama bulan Juni 2015. Hal ini dilakukan mengingat beberapa penyedia konten memiliki perbedaan dalam mendefinisikan kategori berita pada kontennya, sehingga dapat diketahui hasil Kluster yang akurat tanpa memperhatikan sumber kontennya. Beberapa penyedia konten akan dijadikan sebagai sumber data adalah sebagai berikut :

1. <http://detik.com>
2. <http://www.kompas.com>
3. <http://www.liputan6.com>

3.3 Variabel Penelitian

Variabel-variabel pada penelitian ini adalah sebagai berikut :

- Rasio Energi yang Tersisa : merupakan variabel numerik yang menggambarkan rasio antara norm nilai singular asli dengan nilai singular yang tereduksi.
- *Recall* : merupakan ukuran yang menyatakan rasio jumlah kemunculan konten yang tepat pada suatu Kluster, dengan jumlah keseluruhan konten yang seharusnya masuk dalam Kluster tersebut.
- Presisi : merupakan ukuran yang menyatakan rasio jumlah kemunculan konten yang tepat pada suatu Kluster, dengan jumlah keseluruhan konten yang terdapat pada Kluster tersebut.
- *F-Measure* : Merupakan rerata harmonik antara presisi dan *recall*.

3.4 Hipotesis

Statemen hipotesis yang dapat ditarik pada penelitian ini adalah “pendekatan Kluster berdasarkan ciri *TF/IDF* yang tereduksi memiliki tingkat akurasi lebih baik jika dibandingkan dengan pendekatan berdasarkan ciri *TF/IDF* asli”. Kualitas masing-masing ciri digambarkan melalui ukuran *recall*, presisi dan *F-Measure*.

