

## **BAB 3**

### **METODOLOGI PENELITIAN**

#### **3.1 Metodologi Penelitian**

Pada penelitian ini diterapkan metodologi yang terdiri dari beberapa tahap, yakni sebagai berikut:

1. Telaah Literatur

Pada tahap telaah literatur, yang dilakukan adalah pencarian dan pemahaman teori-teori serta metode-metode yang akan digunakan dalam penelitian ini, seperti teori terkait dengan *topic modeling* dan LDA. Informasi yang didapatkan untuk tahap telaah literatur ini berasal dari berbagai jurnal penelitian dan karya tulis ilmiah yang terkait dengan topik penelitian ini.

2. Perancangan Aplikasi

Pada tahap ini dibutuhkan perancangan metode LDA yang akan diimplementasikan. Alur algoritma yang diimplementasikan akan dijelaskan menggunakan *flowchart*. Perancangan kode untuk mengimplementasikan LDA juga akan dilakukan.

3. Pengumpulan Data

Pengumpulan set data akan didapatkan dari *tweet* pada sosial media Twitter. Data yang dikumpulkan berjumlah 50000 *tweet* didapatkan dari

pengumpulan data sejumlah 5000 *tweet* per hari dalam rentang waktu sepuluh hari.

#### 4. Implementasi

Pada tahap implementasi, penulisan kode implementasi metode LDA untuk melakukan *topic modeling* dilakukan. Pada tahap ini juga dilakukan penulisan kode untuk proses *preprocessing* sebagai tahapan awal sebelum dilakukan proses *topic modeling*.

#### 5. Uji Coba dan Debugging

Tahap uji coba bertujuan untuk menguji apakah metode LDA dapat *diimplementasikan* dan berjalan dengan baik dan juga sebagai pengukuran performa dengan menggunakan *topic coherence* dari hasil *topic modeling* dengan metode LDA. Selain itu, *debugging* pada kode yang telah diimplementasikan juga akan dilakukan untuk mengetahui serta memperbaiki kesalahan dan kekurangan dari fungsionalitas sistem yang akan dikembangkan.

#### 6. Dokumentasi

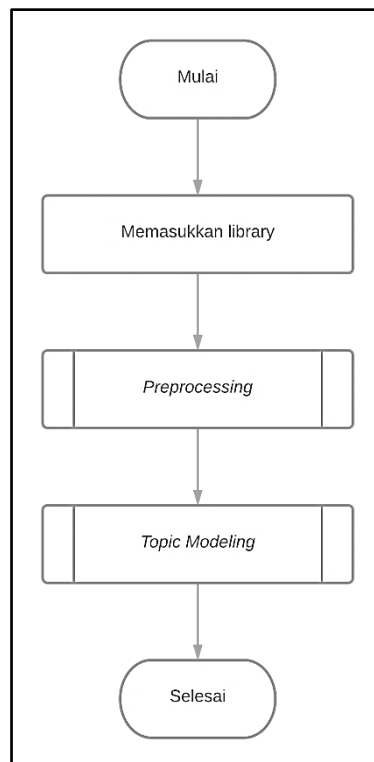
Dokumentasi *yang* akan dilakukan adalah dengan penulisan laporan dari penelitian dan implementasi yang dilakukan. Laporan akan dibuat mulai dari pendahuluan hingga simpulan dan saran.

## 3.2 Perancangan Aplikasi

Perancangan aplikasi pada penelitian ini direpresentasikan menggunakan *flowchart*. Terdapat *flowchart* penelitian secara umum yang berisikan kerja dari sistem secara keseluruhan, kemudian terdapat juga *flowchart* penjelasan lebih lanjut dari perancangan sistem yang dilakukan.

### 3.2.1 Flowchart Penelitian Secara Umum

Gambar 3.1 adalah *flowchart* atau alur kerja dari penelitian secara umum yang merupakan proses *topic modeling* menggunakan metode LDA. Alur kerja penelitian secara umum ini terbagi menjadi dua bagian besar yaitu proses *preprocessing* dan proses *topic modeling*. Pada penelitian ini, alur kerja tersebut akan dilakukan dua kali secara terpisah dengan dua set data yang berbeda.



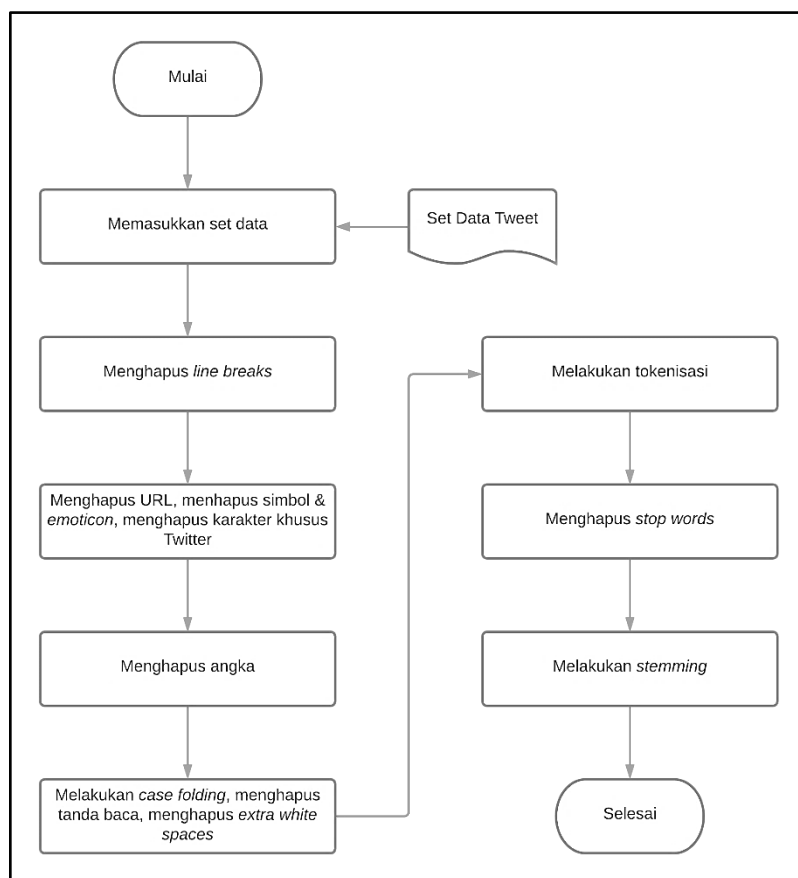
Gambar 3.1 Flowchart penelitian secara umum

Sebagai awal, dilakukan proses memasukkan *library* yang dibutuhkan untuk membantu proses penelitian. Selanjutnya sistem melakukan proses *preprocessing*

untuk mendapatkan bentuk set data yang diinginkan. Kemudian hasil *preprocessing* tersebut digunakan untuk proses *topic modeling* untuk mendapatkan ragam topik beserta hasil evaluasi berupa *topic coherence* dari *topic modeling* menggunakan metode LDA.

### 3.2.2 Flowchart Preprocessing

Proses *preprocessing* ini menggunakan set data yang telah dikumpulkan sebelumnya. Gambar 3.2 adalah *flowchart* atau alur kerja dari proses *preprocessing*. *Preprocessing* dilakukan dengan tujuan mengolah set data terlebih dahulu untuk mendapatkan bentuk set data yang diinginkan sebelum diolah sesuai tujuan penelitian.



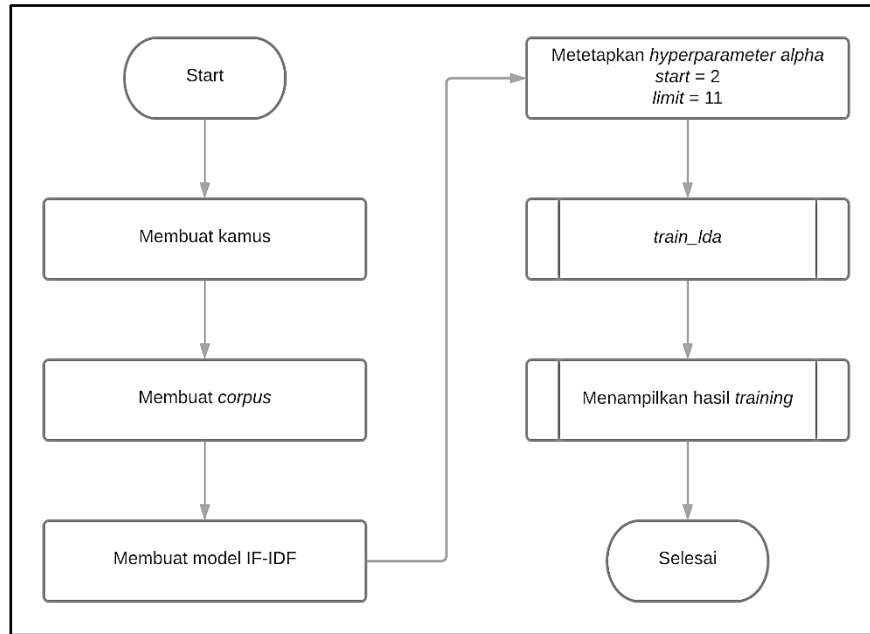
Gambar 3.2 Flowchart preprocessing

### 3.2.3 Flowchart Proses Topic Modeling

Setelah dilakukannya proses *preprocessing*, dilakukan proses *topic modeling*. Proses ini menggunakan bantuan *library* dari Gensim. Proses pertama yang dilakukan adalah membentuk sebuah *dictionary* atau kamus dari set data yang telah diolah sebelumnya. Kemudian set data dan kamus diolah kembali menjadi sebuah *corpus* dalam bentuk atau model *Bag of Words* (BoW). Selanjutnya dilakukan metode *Term Frequency - Inverse Document Frequency* (TF-IDF) pada *corpus* sebelumnya.

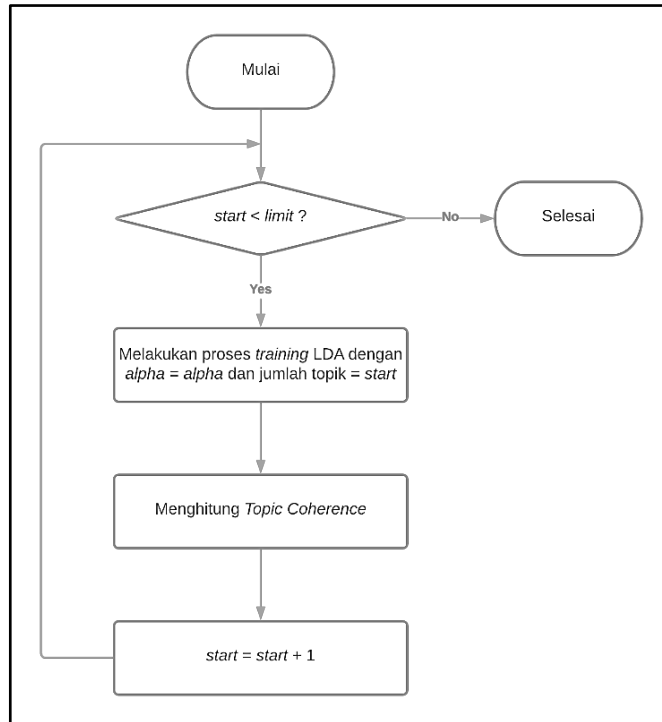
Setelah *corpus* sudah siap, maka dilakukan penetapan *hyperparameter alpha* dan *hyperparameter* jumlah topik yang telah ditentukan yaitu 2, 3, 4, 5, 6, 7, 8, 9, 10. Kemudian dilanjutkan subproses *train\_lda* yang merupakan proses *training corpus* untuk melakukan *topic modeling* menggunakan metode LDA. Pada subproses *train\_lda* ini juga dilakukan perhitungan *topic coherence* sebagai evaluasi nantinya. Setelah itu dilakukan subproses menampilkan hasil *training* yang di dalamnya dilakukan proses menampilkan grafik *coherence score*, menampilkan nilai *coherence score*, memilih *hyperparameter* jumlah topik terbaik, dan menampilkan sepuluh kata yang memiliki kontribusi paling baik pada setiap topik.

Pada penelitian ini, proses penetapan *hyperparameter alpha* hingga subproses menampilkan hasil *training* tersebut dilakukan berulang sebanyak lima kali dengan *hyperparameter alpha* yang berbeda, yaitu 0.1, 0.25, 0.5, 0.75, dan 1. *Flowchart* dari proses *topic modeling* dapat dilihat pada Gambar 3.3.

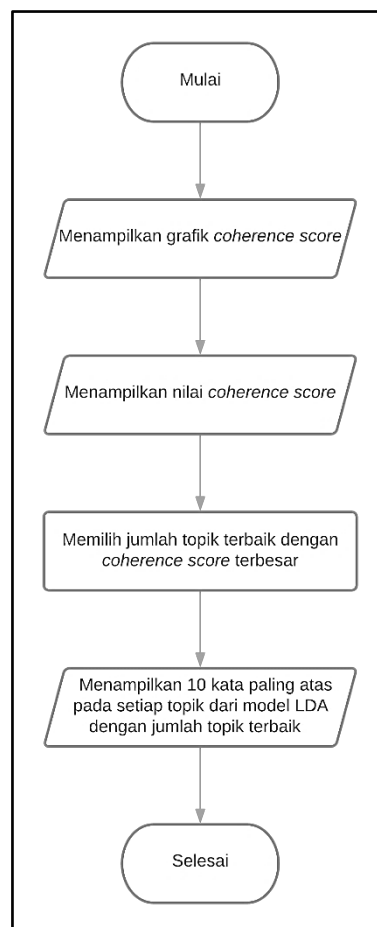


Gambar 3.3 Flowchart proses topic modeling

Pada *flowchart* proses *topic modelling* terdapat subproses yaitu subproses *train\_lda* dan subproses menampilkan hasil *training*. Pada subproses *train\_lda* dilakukan proses *training* model LDA beserta perhitungan *topic coherence* untuk membantu pengukuran performa. *Flowchart* subproses *train\_lda* dapat dilihat pada Gambar 3.4. Gambar 3.5 merupakan *flowchart* subproses menampilkan hasil *training* untuk membantu melihat performa dari hasil *training*. Pada subproses ini, ditampilkan visualisasi berupa grafik yaitu *coherence score* dari setiap *hyperparameter* jumlah topik. Setelah itu dilanjutkan dengan menampilkan *coherence score* dari setiap *hyperparameter* jumlah topik dan memilih *coherence score* paling besar. Kemudian ditampilkan juga sepuluh kata paling atas dengan nilai yang dimilikinya pada setiap topik dari hasil *training corpus* dengan *hyperparameter* jumlah topik dengan *coherence score* paling besar dari hasil yang telah ditentukan.



Gambar 3.4 Flowchart subproses train\_lda



Gambar 3.5 Flowchart subproses menampilkan hasil training