# **BAB III**

**METODELOGI PENELITIAN**

**3.1 Objek Penelitian**

Penelitian “Pengelompokan Judul Penelitian Dosen Menggunakan Metode *Suffix Tree Clustering*” mengambil obyek penelitian dari Lembaga Penelitian dan Pengembangan Universitas Ahmad Dahlan berupa data judul Penelitian Dosen UAD dari tahun 2015 hingga tahun 2017 dengan atribut Peneliti Utama, Prodi, Fakultas, Judul Penelitian, dan Tahun dalam bentuk *file excel*.

**3.2 Metode Pengumpulan Data**

**3.2.1 Wawancara**

Wawancara dilakukan dengan cara tanya jawab secara langsung dengan Drh. Asep Rustiawan, M.S selaku Sekretaris Lembaga Penelitian dan Pengembangan Universitas Ahmad Dahlan terkait pengelompokan yang dilakukan oleh pihak LPP UAD.

**3.2.2 Data *Private***

Metode ini dimaksudkan untuk mendapatkan data *private* berupa judul penelitian dosen Universitas Ahmad Dahlan berupa *file excel* dari Sekretaris LPP UAD.

**3.2.3 Studi Literatur**

Pengumpulan data dilakukan dengan studi literatur dari berbagai macam buku, artikel, publikasi ilmiah dan *e-book* untuk mempelajari :

1. Mekanisme *text mining* dalam mengelompokan data.
2. Mekanisme pengelompokan data menggunakan metode *Suffix Tree Clustering.*

**3.3 Alat Penelitian**

Alat penelitian dibagi menjadi dua, yaitu perangkat keras dan perangkat lunak. Alat penelitian yang dibutuhkan dalam melakukan penelitian ini adalah sebagai berikut:

* + 1. **Kebutuhan Perangkat Keras (Hardware)**

1. Seperangkat Laptop
   * 1. **Kebutuhan Perangkat Lunak (Software)**
2. Editor Bahasa pemrograman *Python*, *Microframework Flask,* dan *Sublime Text* 3.
3. Database Server yang digunakan yaitu MongoDB dan Database Client yang digunakan yaitu Robomongo.
4. Web Browser seperti Google Chrome, Mozilla Firefox, Internet Explorer, dan sebagainya.
5. Modul bahasa pemrograman Python seperti Anaconda
6. Library Pemrograman yang digunakan *suffix\_trees*, python *suffix tree*, pysuffix dan sebagainya.

**3.4 Tahapan Text Mining**

Tahapan *text mining* yang akan diterapkan pada penelitian ini adalah dimulai dengan *Load* Data teks, lalu data teks masuk ke tahapan text mining yaitu *tokenizing, fitering, stemming.* Kemudian pemahaman dari Algoritma *Suffix Tree Clustering,* dan perhitungan Document Frequency & Term Frequencydilanjutkan dengan implementasi algoritma tersebut ke dalam pengelompokan Judul Penelitian Dosen. Adapun *Flowchart* Alur Proses *Clustering* Judul Penelitian Dosen sebagai berikut:

Load Data

Tokenizing

Filtering

Stemming

Algoritma *Suffix Tree Clustering*

Menghitung Term Frequency & Inverse Document Frequency

Clustering Judul Penelitian Dosen

Pengujian

Gambar 3.1. Tahapan Text Mining.

**3.4.1 Load Data**

Load Data merupakan awal dari tahapan text mining. Judul penelitian dosen yang telah dikumpulkan dalam bentuk *excel* kemudian di *load* kedalam program agar judul penelitian dosen dapat diolah pada tahapan selanjutnya.

**3.4.2 *Tokenizing***

Setelah judul penelitian berhasil diupload maka tahapan selanjutnya adalah tahapan *tokenizing* yaitu tahap pemotongan *string* berdasarkan tiap kata penyusunnya. Judul penelitian dosen akan dipecah menjadi perkata. Contoh : Instrumen Pengukuran Kesadaran Kanker Payudara di Yogyakarta, setelah di *tokenizing* menjadi “Instrumen”, “Pengukuran”, “Kesadaran”, “Kanker”, “Payudara”, “di”, “Yogyakarta”.

**3.4.3 *Filtering***

*Filtering* merupakan proses pengecekan satu persatu kata-kata yang ada pada setiap judul penelitian dosen kemudian membuang kata-kata yang dianggap tidak penting atau yang biasa disebut dengan *stoplist/stopword*. Kata yang dihapus adalah kata hubung atau kongjungsi yaitu “yang”, “dan”, ”di”, “untuk”, dan lain sebagainya. Contoh : “Instrumen”, “Pengukuran”, “Kesadaran”, “Kanker”, “Payudara”, “di”, “Yogyakarta”, setelah di *filtering* menjadi “Intrumen”, “Pengukuran”, “Kesadaran”, “Kanker”, “Payudara”, “Yogyakarta”.

**3.4.3 *Stemming***

*Stemming* merupakan proses penghilangan / pemotongan *prefix* (awalan) dan *sufiks* (akhiran) dari kata dan istilah-istilah dokumen. Dalam data judul skirpsi, *stemming* digunakan untuk penghilangan / pemotongan awalan dan akhiran dari kata-kata yang sudah melalui tahapan *tokenizing* dan *filtering*. Contoh : “Intrumen”, “Pengukuran”, “Kesadaran”, “Kanker”, “Payudara”, “Yogyakarta” setelah melalu tahap *stemming* menjadi “Instrumen”, “Ukur”, “Sadar”, “Kanker”, “Payudara”, “Yogyakarta”.

**3.4.4 Algoritma *Suffix Tree Clustering***

Algoritma *Suffix Tree Clustering* merupakan suatu hasil pencarian yang menerapkan penggunaan algoritma *clustering.* Algoritma *Suffix Tree Clustering* (STC) memiliki dua kunci utama, yaitu:

1. Menggunakan *phrase* sebagai dasar pembentukan *clusternya.*
2. Menggunakan suatu definisi *cluster* sederhana.

*Suffix Tree Clustering* memiliki dua langkah utama. Dalam langkah pertama, pencarian *shared phrase* untuk semua dokumen berita yang dikoleksi. *Shared phrase* disebut sebagai *phrase cluster* atau *base cluster,* yang ditemukan dengan menggunakan suatu struktur data. Dalam langkah kedua, mengkombinasikan *base cluster* – *base cluster* ke dalam suatu *cluster.* Penggabungan antara dua *base cluster* didasarkan pada jumlah dokumen yang melakukan *overlap* diantara kedua *base cluster* tersebut. Suatu *phrase* yang dimaksud dalam konteks algoritma ini adalah urutan satu atau lebih kata-kata. STC memiliki tiga langkah utama, yaitu :

1. *Cleaning Dokumen.*
2. Identifikasi *Base Cluster* menggunakan *Suffix Tree.*
3. Mengkombinasikan *Base Cluster* kedalam suatu *cluster.*

**3.4.5 *Term Frequency &* Inverse *Document Frequency***

*Term Frequency & Inverse Document Frequency* (TF-IDF) merupakan metode untuk menghitung bobot. Metode TF-IDF terkenal *efisien,* mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency (TF) dan Inverse Document Frequency (IDF)* pada setiap kata disetiap dokumen dalam korpus. Metode ini akan menghitung bobot setiap kata didalam dokumen.

**3.4.6 *Clustering* Judul Penelitian Dosen**

*Clustering* Judul Penelitian Dosen merupakan hasil akhir dari proses *tokenizing, filtering, stemming,* perhitungan menggunakan metode *Term Frequency & Inverse Document Frequency* dan pengelompokan menggunakan algoritma *Suffix Tree Clustering.* Visualisasi *clustering* judul penelitian dosen berupa kelompok judul penelitian dosen yang telah ditentukan berdasarkan bidang penelitiannya dan berupa pohon yang akan memvisualisasikan judul penelitian dosen pada setiap bidang penelitiannya.

**3.4.7 Pengujian**

Pengujian akan dilakukan untuk mengetahui sebaik apakah algoritma *suffix tree clustering* dalam hal akurasi. Pada pengujian dilakukan menggunakan perhitungan validitas *Term Frequency & Inverse Document Frequency* (TF-IDF).

**3.5 Implementasi**

Implementasi merupakan tahapan untuk mengimplementasikan perancangan dengan menggunakan algoritma *Suffix Tree Clustering* ke dalam system. Dengan menggunakan bahasa pemrograman *Phyton* akan menghasilkan pengelompokan judul penelitian dosen. Tahap implementasi menerapkan tahapan *text mining* pada system.

**3.6 Pengujian**

Setelah system selesai dibuat maka tahapan selanjutnya adalah pengujian terhadap sistem. Pengujian ini bertujuan untuk menguji apakah system sudah berjalan sesuai keinginan atau belum. Pengujian terhadap system dilakukan dengan metode *Suffix Tree* untuk *clustering.* Tahap pengujian dengan *Suffix Tree* untuk mengetahui seberapa akurat yang dihasilkan dari pengelompokan Judul Penelitian Dosen menggunakan metode *Suffix Tree Clustering.*

E. Jadwal Penelitian

Jadwal pelaksanaan penelitian selama 6 bulan mulai dari bulan Januari 2019 sampai bulan Juni 2019 dengan rincian seperti table berikut ini:

**Tabel 1. Rencana Jadwal Penelitian**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Kegiatan** | **Bulanan Ke -** | | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** |
| 1 | Seminar Proposal |  |  |  |  |  |  |
| 2 | Pengumpulan Data |  |  |  |  |  |  |
| 3 | Pengolahaan Data dalam proses Text Mining ( Tokenizing, Filtering, Stemming, Tagging, Analizing) |  |  |  |  |  |  |
| 4 | Penerapan Algoritma *Suffix Tree Clustering* |  |  |  |  |  |  |
| 5 | Implementasi Sistem |  |  |  |  |  |  |
| 6 | Pengujian Sistem |  |  |  |  |  |  |
| 7 | Evaluasi Sistem |  |  |  |  |  |  |
| 8 | Penyelesaian Laporan |  |  |  |  |  |  |
| 9 | Pendadaran |  |  |  |  |  |  |