**BAB III**

**METODE PENELITIAN**

**A. Desain Penelitian**

Metode penelitian yang digunakan adalah metode eksperimen. Metode eksperimen adalah metode cara penyajian sebuah data melalui proses percobaan-percobaan untuk membuktikan sendiri suatu pernyataan atau hipotesis tertentu dalam sebuah penelitian.Penelitian eksperimen merupakan penelitian yang paling valid dikarenakan metode eksperimen dilakukan dengan pengontrolan variable yang sangat ketat sehingga didapat hasil yang maksimal(Jaedun, 2011). Keuntungan utama penelitian dengan penggunaan metode eksperimen adalah adanya kendali ditangan peneliti dan ketepatan sebuah logika yang terkandung di dalam metode tersebut. Metode ini juga sering digunakan karena dapat dibuat perencanaan yang sistematis sehingga metode eksperimen adalah metode yang paling kuat dalam sebuah penelitian. Berikut adalah tahapan metode eksperimen :

Pengumpulan Data

Pre Prosses NLP

Proses TF-IDF

Proses Apriori

Experimen dan Pengujian

Evaluasi dan Validasi Hasil

**Gambar 3.1 Metode yang Diusulkan**

1. **Identifikasi Berita *Hoax***

Identifikasi berita *hoax* dilakukan dengan cara mencari kata – kata yang sering atau yang memungkinkan untuk digunakan sebagai kata yang terindikasi sebagai kata hoax. Seperti contoh , kata “akan”. Kata “akan” biasa digunakan untuk sebuah premis yang belum dilakukan. Sedangkan berita adalah sebuah informasi yang faktual bersifat sedang atau sudah terjadi. Sehingga kata “akan” memiliki peluang sebagai kata yang terindikasi *hoax*.Pada *system* deteksi *hoax* ini, dilakukan pencarian kata yang memungkinkan untuk pembuatan berita hoax dengan menggunakan rumus pembobotan tf-idf dan apriori sehingga didapat nilai sebuah kata berkemungkinan dijadikan untuk berita *hoax.*

**2. Pengumpulan Data**

Data yang dikumpulkan terdiri dari 2 buah jenis data yakni data uji dan data latih. Data latih yakni data yang berisikan kumpulan berita *hoax* yang nantinya akan digunakan untuk proses pengindex an kata guna pembuatan *library*. Data uji yakni data yang berisikan berita baik *hoax* maupun real yang nantinya digunakan untuk menentukan tingkat akurasi sistem. Data yang diambil adalah data berita *hoax* dari situs turbackhoax.id dan juga data *hoax* dari broadcast whatsapp. Sedangkan berita yang nantinya dijadikan data uji berasal dari detik.com dan sebagian berasal dari turnbackhoax.id.

**3. Pre Prossesing Menggunakan NLP (Neural Language Processing)**

Data berita yang didapat dari situs turnbackhoax.id ataupun dari detik.com adalah sebuah data mentah yang belum dapat diolah. Agar data tersebut dapat diolah, maka dilakukan pre prosessing dengan menggunakan NLP (*Neural Language Processing*). Pada tahap pre prosessing ini memiliki beberapa *step* yakni :

3.1. Tokenizing

Proses pemisahan text menjadi per kata sehingga tiap kata dapat diolah dengan mudah. Proses pemisahan kata dilakukan menggunakan fungsi dalam python.

3.2. *Stoward Removal*

Stopword merupakan kata umum yang digunakan / kata yang tidak penting untuk dianalisi / diproses. Seperti contoh kata : dan, atau, tetapi. Tujuan dari penghapusan kata ini yakni mengurangi pembengkakan jumlah index yang digunakan.

Cek pada library kata umum

Terdapat Kata Umum

Hapus Kata

*Input* Data

Yes

No

**Gambar 3.2 *Flowchart Stoward Removal***

a. Masukkan Data

b. Proses pencarian kata yang dinilai sebagai kata tidak penting pada *library*

c. Jika terdapat kata umum maka *system* akan menghapus kata tersebut

d. Jika tidak, maka *system* akan melanjutkan pencarian sampai akhir dari berita

f. *System* mengakhiri proses

3.3 *Stemming*

Sebuah penghapusan kata yang memilki awalan / akhiran sehingga didapat kata dasar. Guna dari *stemming* ini yakni meminimalisir persamaan kata yang memiliki perbedaan awalan / akhiran sehingga dapat memperkecil jumlah indeks yang akan diproses. Proses stemming ini menggunakan fungsi *stemmer factory* yang terdapat pada python dengan *import* *package* sastrawi

**4. *Processing* menggunakan Tf-Idf**

Setelah dilakukan *pre processing* dengan menggunakan NLP, maka langkah selanjutnya yakni menacari bobot dari sebuah kata dengan menggunakan *Term Frequent dan Inverse Document Frequent.* Alur sistem dari program Tf-Idf ini digambarkan pada *flowchart* berikut :

Hitung TF

Tf=Frekuensi kemunculan kata /seluruh kata

Hitung IDF Idf=log(jml seluruh doc/jml doc mengandung kata)

Data NLP

Bobot = Tf\*Idf

Input into csv

Kata+bobot

**Gambar 3.3 *Flowchart* Tf-Idf**

1. Data yang sudah di proses dengan NLP dimasukkan kedalam perhitungan Tf
2. Selanjutnya akan dihitung nilai Idf dari data
3. Pencarian bobot dengan menggabungkkan Tf dan Idf
4. Kata dan bobot di simpan dalam bentuk .csv

**5. *Processing* menggunakan Apriori**

Tahap selanjutnya yakni prosesing dengan menggunakan metode apriori. Dalam metode ini diambil sebuah nilai support dari perbandingan antara jumlah berita yang mengandung kata dengan jumlah total berita. Yang selanjutnya akan dimasukkan kedalam data *index library* kata yang tersimpan beserta bobot. Selanjunya, data akan dicari rataan dari nilai bobot dan nilai support sehingga didapat nilai akhir. Proses alur dari sistem digambarkan pada *flowchart* berikut :

Hitung Support

Support=jumlah berita yang mengandung kata /seluruh berita

Hitung Nilai Akhir : (Support+Bobot) / 2

Data NLP

Input into csv

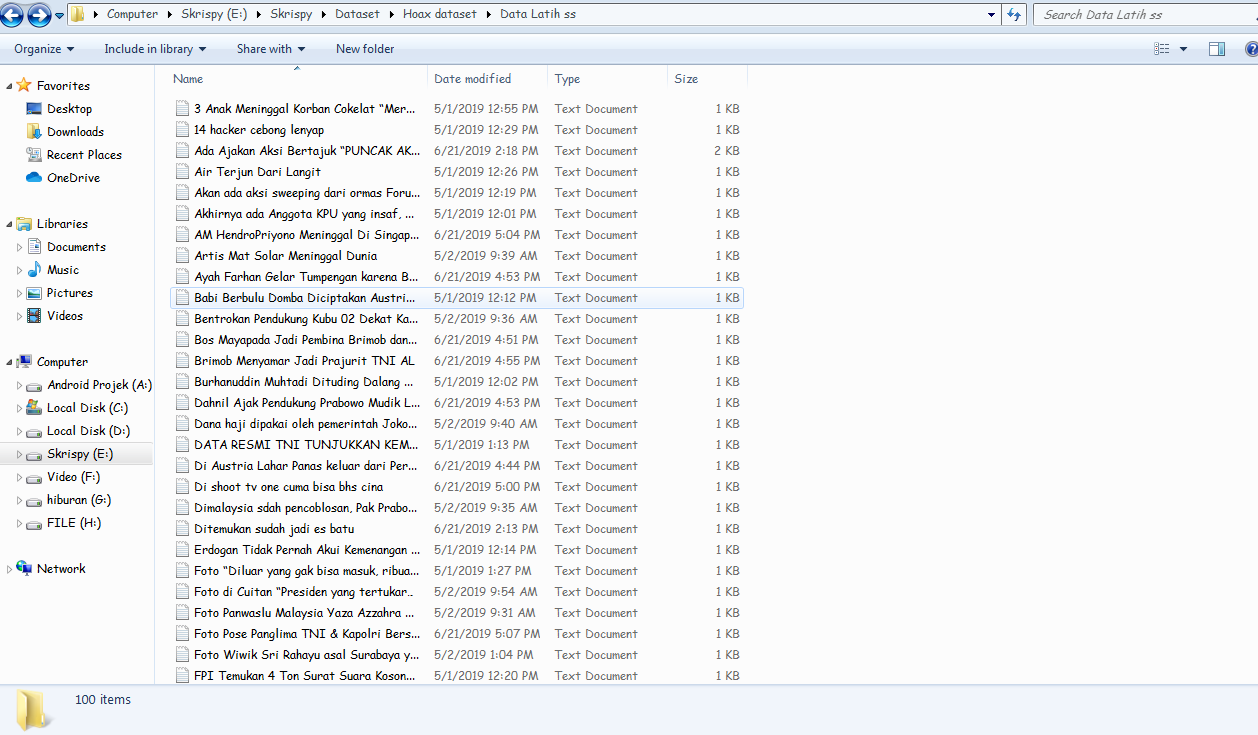
Kata+Support+Nilai Akhir

**Gambar 3.4 *Flowchart* *Apriori***

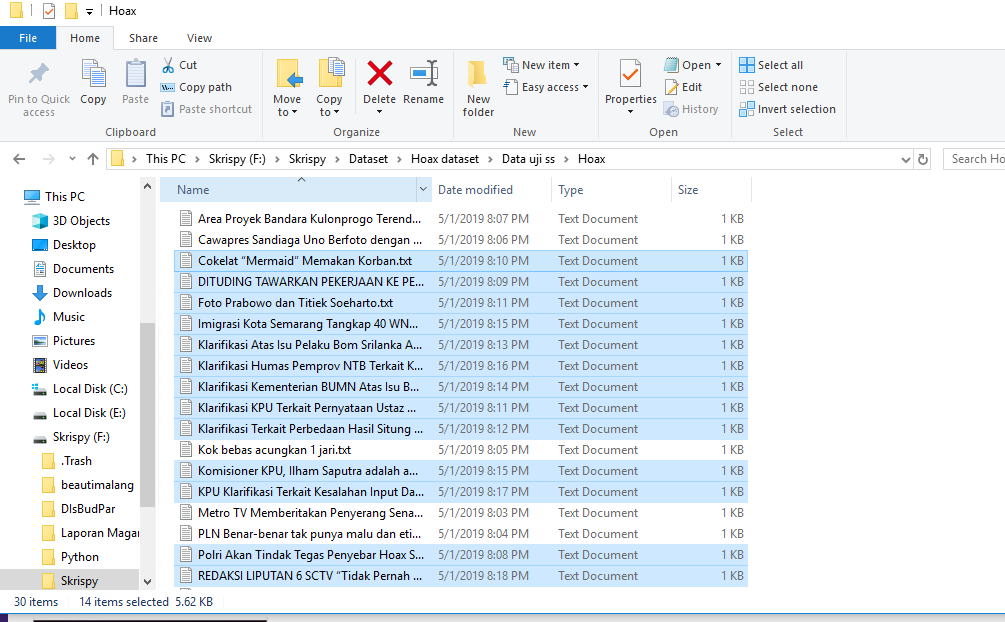
1. Masukan data yang sudah diproses NLP
2. *System* menghitung nilai *support* dengan menggunakan rumus apriori
3. Proses selanjutnya yakni pencarian nilai akhir dengan cara menari rataan dari nilai *support*  dan nilai bobot
4. Data kata dan nilai akhir di simpan dalam bentuk .csv

**6. Eksperimen dan Pengujian**

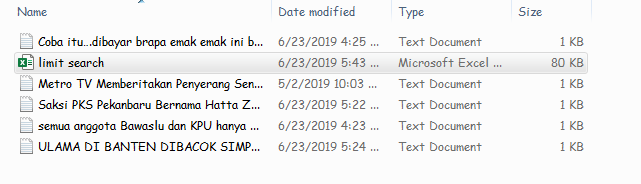
Penelitian ini akan menggunakan data latih sebanyak 100 buah berita *hoax* yang dapat di lihat pada Lampiran 1. Dan 30 buah data uji dengan rincian 16 buah berita hoax dan 14 buah berita real yang dapat dilihat pada Lampiran 2. Dengan data uji tersebut dilakukan pengujian sistem untuk mencari tingkat akurasi dari *sytem*.Untuk pencarian nilai limit akan digunakan 5 buah berita *hoax* yang nantinya akan diproses sehingga didapat batas yang sesuai untuk dijadikan acuan.Data Uji dan data latih dapat dilihat pada Gambar 3.5 dan 3.6. Pada Gambar 3.5 terdapat 100 buah data latih berita *hoax* yang nantinya akan diproses oleh *system* sehingga dapat menghasilkan *library* kata *hoax.* Pada Gambar 3.6 yakni terdapat 30 buah berita dengan rincian : 16 buah berita hoax dan 14 buah berita real yang nantinya akan digunakan untuk pengujian akurasi *system.*Dan pada Gambar 3.7proses pencarian nilai limit. Terdapat 5 buah berita terindikasi *hoax* yang nantinya akan digunakan sebagai *sample* data guna mencari nilai limit dari system yang digunakan sebagai acuan pengklasifikasian berita.

****

**Gambar 3.5 Data Latih**

****

**Gambar 3.6 Data Uji**



**Gambar 3.7 Data Pencarian Limit**

**7. Evaluasi dan Validasi Hasil**

Evaluasi dilakukan dengan memasukkan hasil dari data uji yang telah diuji pada tahap sebelumnya. Digunakan rumus *precision* *and* *recall* dan *accuracy* untuk menentukan tingkat keakuratan sistem.