

B. Rumusan Masalah

Berdasarkan uraian penulis tentang konteks masalah penelitian, maka penulis membentuk suatu masalah yang berkaitan dengan topik penulisan, yaitu:

1. Bagaimana cara crawling portal berita online ?
2. Bagaimana perbandingan performa model klasifikasi untuk sentimen isi berita ?
3. Bagaimana hasil topik yang dominan dari isi berita menggunakan model topic *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA) ?
4. Bagaimana implementasi dari hasil model sentimen dan pemodelan topik ?

C. Batasan Masalah

Batasan - batasan masalah pada penyusunan penelitian dengan topik penulisan , yaitu:

1. Berita-berita yang dianalisis adalah berita-berita yang diterbitkan oleh media massa Indonesia. Berita Online yang dipakai yaitu Detik, CNN dan Kompas.
2. Pengambilan berita online di lakukan dari 1 Oktober 2023 sampai 31 Desember 2023
3. Analisis isi berita untuk melibatkan sentimen positif, negatif dan netral.
4. Analisis isi berita untuk melihat pemodelan topik yang terkandung dalam berita selama pemilu,
5. Perbandingan algoritme klasifikasi yang digunakan dalam penelitian ada 4 algoritme klasifikasi yaitu Naiye Bayes, *Support Vector Machine* (SVM), Logistic Regression , dan Random Forest

6. pemodelan topik *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA) menggunakan nilai *coherence score* mendapatkan nilai jumlah topik.
7. Filter kata yang terkait pemilu untuk dataset pemilu yaitu pemilu, mk, kpu, ganjar, prabowo, anies, cak imin, gibran, mahfud md, capres, cawapres, bawaslu, dan partai.

D. Tujuan Penelitian

Tujuan yang hendak dicapai yaitu menyelesaikan masalah yang diangkat di atas secara spesifik, yaitu:

1. Memaparkan cara crawling portal berita online .
2. Mengetahui model algoritme klasifikasi yang memiliki akurasi yang paling baik untuk sentimen isi berita pemilu 2024.
3. Mengetahui hasil topik yang dominan dari isi berita dari berita pemilu 2024.
4. Implementasi dari hasil model sentimen dan pemodelan topik.

E. Metode Penelitian

Pada bagian ini akan menjelaskan tentang metodologi penelitian yang akan digunakan. Berupa pengumpulan data, *preprocessing* data, klasifikasi data sentimen pada isi berita dengan menggunakan dengan perbandingan beberapa algoritme klasifikasi, pencarian topik pada isi berita dalam bentuk *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA). Hasil akhir yang mendapat model yang terbaik dan mendapat topik dari LSA dan LSA

1. Pengumpulan Data

Tahap pertama membuat crawling berita dimana tujuan mengambil data yaitu judul berita, isi berita , tanggal

berita, link berita dan asal berita dimana portal berita yang diambil 3 yaitu Detik, CNN dan Kompas dimana membuat service pengumpulan data dengan menggunakan library python yaitu beautifulsoup4 dan request. Dalam pengambilan data akan dimasukkan ke dalam database sqlite.

2. *Preprocessing Data*

Pada tahap ini akan memproses data yang sudah diambil dari berita tadi. Data yang diambil sebelumnya masih tidak mempunyai struktur maka dari itu proses ini akan membuat data menjadi terstruktur. Preprocessing mempunyai beberapa tahap, yaitu stopwords, tokenizing, dan case folding. preprocessing yang dilakukan akan menggunakan jupyter notebook.

3. Klasifikasi Data Sentimen Pada Isi Berita dengan Menggunakan Perbandingan Beberapa algoritme

Pada tahap ini melakukan pelabelan secara manual bertujuan untuk memberi bobot pada data kepada isi berita dimana dibagi kedalam tiga nilai yaitu positif dengan nilai 1 lalu netral poin 0 dan negatif dengan nilai -1 dan juga melakukan pelabelan otomatis dengan model dengan hasil output sentimen label dan skor sentimen di setiap isi berita yang diinputkan. Sentimen yang sudah dilabeli secara manual dan otomatis akan di masukan kedalam berapa model untuk melihat nilai akurasi dan F1 skor untuk dan di buat perbandingan antara model klasifikasi untuk mencari model yang paling baik dari segi nilai akurasi dan F1 skor.

4. Pencarian Topik Pada Isi Berita dalam Bentuk *Latent Dirichlet Allocation (LDA)* dan *Latent Semantic Analysis (LSA)*

Tahap ini mengolah data yaitu isi berita yang dimana

melihat isi berita dengan jumlah topik dan top kata nya yang sering muncul dengan metode LDA dan LSA dengan menentukan jumlah topik yang di butuhkan menggunakan nilai *coherence score*.

5. Hasil Sentimen dan Topik *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA)

Tahap ini menyimpulkan pada algoritme klasifikasi apa yang paling baik di gunakan. LDA dan LSA membahas hasil jumlah topik yang di gunakan dan membahas top kata dari setiap topik.

F. Sistematika Penulisan

Berikut ini merupakan sistematika dalam penulisan yang akan dilakukan pada penelitian ini:

BAB I PENDAHULUAN

Bab 1 merupakan pendahuluan, yang berisi tentang penjelasan mengenai latar belakang masalah yang diambil untuk penelitian, rumusan masalah, Batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan yang dikerjakan.

BAB II TINJAUAN PUSTAKA

Bab 2 merupakan tinjauan pustaka,berisi mengenai perbandingan penelitian-penelitian yang sudah ada sebelumnya. Penelitian yang dibandingkan juga relevan dengan penelitian yang sedang diambil.

BAB III LANDASAN TEORI

Bab 3 merupakan landasan teori yang menjelaskan mengenai teori-teori yang digunakan dan berkaitan dalam penelitian. Seperti crawling data, analisis sentimen, penjelasannya masing algoritme klasifikasi yang digunakan penelitian ini, *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA)

BAB IV METODOLOGI

Bab 4 merupakan metodologi, menjelaskan tentang metodologi

yang digunakan untuk penelitian. Metodologi tersebut seperti crawling berita, preprocessing data, klasifikasi sentimen pada isi dengan perbandingan beberapa algoritme klasifikasi , membuat model topik pada isi berita dengan metode *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA) , Hasil akhirnya adalah hasil sentimen pada isi berita dan hasil topik LDA dan LSA

BAB V HASIL DAN PEMBAHASAN

Bab 5 yang merupakan hasil dan pembahasan, berisi mengenai hasil dari penelitian serta pembahasan yang relevan terhadap penelitian.

BAB VI KESIMPULAN DAN SARAN

Bab 6 yang merupakan kesimpulan dan saran, berisi mengenai kesimpulan dari keseluruhan penelitian yang dilakukan.

