**ABSTRAK**

Nama : Ridho Rachmat Giffary

Program Studi : Teknik Informatika

Judul : Analisis Sentimen dan Topic Modeling Mengenai UU ITE pada Media Sosial Twitter Bahasa Indonesia

Sebanyak 202,6 juta penduduk Indonesia menjadi pengguna internet per tahun 2021, Perkembangan internet di Indonesia telah berkembang sangat pesat dalam banyak hal termasuk media dan komunikasi. Perkembangan ini juga membawa pengaruh buruk bagi penggunanya, karna hal tersebut dibuatlah Undang – Undang Informasi dan Transaksi Elektronik atau biasa disingkat dengan UU ITE. Namun pembuatan kebijakan tersebut menuai pro dan kontra dikalangan masyarakat. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat dan topik yang dibicarakan mengenai kebijakan UU ITE. Penelitian ini menggabungkan Analisis Sentiment Lexicon-Based dan Pemodelan Topik metode *Latent-Dirichlet Allocation* (LDA).

**Kata Kunci** : *Analisis Sentimen, Pemodelan Topik, Twitter, UU ITE, Lexicon-Based, LDA, Corpus, Vader*

**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Di zaman modern ini, perkembangan teknologi bertumbuh sangat pesat. Sebanyak 202,6 juta penduduk Indonesia menjadi pengguna internet per tahun 2021, hal ini membuat Indonesia menempati peringkat ketiga se Asia Pasifik sebagai pengguna internet terbanyak [1]. Perkembangan internet di Indonesia telah berkembang sangat pesat dalam banyak hal termasuk media dan komunikasi. Hal ini dapat dilihat dari banyaknya penggunaan media sosial di lingkungan masyarakat. Di masyarakat, media sosial digunakan untuk banyak hal seperti berjualan produk, berkomunikasi dengan keluarga dan teman, mencari lowongan pekerjaan, mencari sumber berita dan informasi, memberikan opini akan suatu hal dan lain sebagainya. Walaupun demikian, perkembangan ini tidak luput dari tindak kejahatan seperti *cyber-bullying*, ujaran kebencian, penyalahgunaan data, berita bohong atau hoaks dan lain sebagainya.

Karna hal tersebut, pemerintah membuat Undang – Undang Informasi dan Transaksi Elektronik atau biasa disingkat dengan UU ITE. Undang – undang ini menuai pro dan kontra diantara kalangan masyarakat, beberapa merasa terbantu karna dapat mengurangi tingkat *cyber-bullying* dan ujaran kebencian di dunia maya, namun banyak juga yang kontra dikarenakan undang – undang ini banyak disalah gunakan oleh pihak yang tidak bertanggung jawab.

Hal ini dapat mempengaruhi emosi pengguna media sosial yang dapat menyebabkan opini sehingga dapat diklasifikasikan untuk menentukan polaritasnya yaitu positif atau negatif. Pada penelitian ini, peneliti menggunakan media sosial twitter untuk mendapatkan data karna twitter merupakan media sosial yang mudah untuk diakses dan banyak penelitian sebelumnya yang menggunakan data dari media sosial ini. Analisis sentiment adalah riset komputasional dari opini sentimen dan emosi yang diekspresikan secara tekstual [2]. Dalam melakukan penarikan data untuk sentimen analisis menggunakan teknik scraping data twitter lalu dilanjut dengan teknik pelebelan data menggunakan *Lexicon-based* lalu dilanjutkan dengan ekstraki fitur menggunakan BoW dan diakhiri dengan topic modeling menggunakan LDA.

* 1. **Rumusan Masalah**

Rumusan masalah yang didapat pada tugas akhir ini adalah:

1. Bagaimana cara melakukan pengambilan serta pengolahan data tweet dari twitter untuk *Lexicon-based* perhitungan Topic Modeling
2. Bagaimana respon masyarakat pengguna media sosial Twitter mengenai kebijakan Undang – Undang Informasi dan Transaksi Elektronik pada tanggal 14 sampai 20 februari 2021
3. Apa saja topik permasalahan yang dibicarakan masyarakat pengguna media sosial Twitter mengenai kebijakan Undang – Undang Informasi dan Transaksi Elektronik pada tanggal 14 sampai 20 februari 2021 dengan kata kunci “UU ITE”
   1. **Batasan Masalah**

Batasan masalah untuk penyusunan tugas akhir ini adalah:

1. Data yang digunakan pada tugas akhir ini adalah data tweet dengan kata kunci “UU ITE” dari Twitter berbahasa Indonesia saja.
2. Pengklasifikasian pada tugas akhir ini menggunakan Teknik *Lexicon-based*
3. Pengelompokan tweet berdasarkan topik pada tugas akhir ini menggunakan metode topik modeling dengan pemodelan LDA
   1. **Tujuan Penelitian**

Tugas akhir ini memiliki tujuan untuk menemukan topik yang sedang dibicarakan mengenai UU ITE pada sosial media Twitter menggunakan pengklasifikasian Lexicon-based dan dilanjutkan dengan topik modeling menggunakan pemodelan LDA.

* 1. **Manfaat Penelitian**

Manfaat yang dapat didapatkan pada penulisan tugas akhir ini adalah:

1. Mendapatkan informasi berupa sentiment positif atau negatif mengenai kebijakan UU ITE
2. Mendapatkan topik yang sedang dibicarakan pengguna media sosial Twitter mengenai kebijakan UU ITE pada tanggal 14 sampai 20 februari 2021
3. Bagi pemerintah, dapat mengetahui sentiment terkait kebijakan UU ITE dan topik yang dibicarakan pengguna media sosial Twitter yang hailnya dapat digunakan untuk bahan evaluasi berdasarkan feedback dari masyarakat

**BAB II**

**TINJAUAN PUSTAKA**

* 1. **Kajian Pustaka**

Penulisan pada penelitian seringkali membutuhkan panduan serta dukungan penelitian lain yang sudah dilakukan sebelumnya yang berkaitan dengan penelitian yang sedang dikerjakan atau berlangsung.

Pada sebuah penelitian yang menganalisi seputar UU ITE menggunakan SVM, mereka menganalisis dengan tahapan crawling data yang berasal dari twitter menggunakan API lalu dilanjutkan dengan tahapan *preprocessing* data, kemudian dilajutkan dengan visualisasi menggunakan *Wordcloud,* lalu melakukan pelabelan sentiment otomatis, kemudian melakukan pembobotan kata dengan TF-IDF dan diakhiri dengan klasifikasi menggunakan metode SVM. Hasil dari penelitian ini menunjukan bahwa sentiment terkait topik yang diambil adalah negative dengan tingkat akurasi 84%[3].

Kemudian pada penelitian lain melakukan menganalisis mengenai Situs Pembajak Artikel Penelitian menggunakan *Lexicon-based,* mereka melakukan analisis dengan melakukan pengumpulan data menggunakan library Python Twint, lalu dilanjutkan dengan proses *preprocessing* untuk membersihkan data yang telah didapat, kemudian dilanjutkan dengan ekstraksi kata kunci, lalu melakukan analisis sentiment. Pada penelitian ini menghasilkan sentimen positif sebesar 55.6%[4].

Lalu pada sebuah penelitian dilakukan penggabungan antara analisis sentiment dan pemodelan topik mengenai BPJS Kesehatan dilakukan pengumpulan data yang bersumber pada media social twitter. Data-data tersebut kemudian dibersihkan pada tahap *preprocessing,* kemudian dilanjutkan dengan tahapan sentiment analisis menggunakan *Vader* dan *topic modeling* menggunakan LDA. Output dari penelitian ini adalah 5 topik mengenai BPJS Kesehatan dan sentiment berupa positive sebesar 62%[5]

Kemudian pada penelitian lain peneliti melakukan gabungan penelitian antara analisis sentimen dan pemodelan topik mengenai produk kosmetik. Penelitian ini melakukan pengumpulan data menggunakan aplikasi *parsehub* pada *market place* di *Female Daily,* lalu dilanjut dengan tahap *preprocessing,* kemudian dilanjutkan dengan implementasi sentiment analisis dan diakhiri dengan topik modeling menggunakan LDA. Keluaran dari penelitian ini adalah 5 topik mengenai produk kosmetik dan sentiment berupa positif[6].

Kemudian pada penelitian lainnya peneliti melakukan gabungan penelitian antara analisis sentimen dan pemodelan topik mengenai aplikasi gojek, peneliti melakukan pengumpulan data dari social media twitter, lalu dilanjutkan dengan proses preprocessing, kemudian dilakukan visualisasi menggunakan *Wordcloud,* lalu dilanjutkan dengan proses pemodelan topik dan diakhiri dengan labeling menggunakan *Vader*. Keluaran dari penelitian ini adalah 5 topik yang yang berhubungan dengan gojek dan sentiment yang didapat adalah positif sebesar 60%[7].

Lalu pada penelitian lain melakukan penelitian gabungan antara analisis sentimen dan pemodelan topik mengenai aplikasi ruangguru. Pada penelitian ini dilakukan pengumpulan data komentar youtube channel ruangguru, lalu dilanjutkan dengan tahapan data cleaning, kemudian dilakukan pelabelan data sentiment dan jenjang Pendidikan, lalu dilakukan visualisasi menggunakan *Wordcloud* berdasarkan sentiment dan jenjang*,* kemudian diakhiri dengan topik modeling berdasarkan jenjang Pendidikan. Output pada penelitian ini adalah 10 topik berdasar jenjang Pendidikan dan sentiment positif[8].

Terakhir, pada penelitian lainnya juga melakukan penelitian gabungan antara analisis sentimen dan pemodelan topik mengenai pembelajaran online. Peneliti melakukan pengumpulan data twitter, lalu dilanjutkan preprocesing data dan kemudian diakhiri dengan sentiment analisis dan topic modeling secara parallel. Keluaran dari penelitian ini adalah 3 topik dan sentiment berupa positif[9].

* 1. **Landasan Teori**
     1. **Analisis Sentimen**

Sentimen analisis merupakan pengelompokkan polaritas dari suatu teks yang berada pada dokumen, kalimat, atau data berbentuk teks dan menentukan apakah opini yang dinyatakan pada dokumen, kalimat atau data berbentuk teks bersifat positive atau negatif. Intinya sentimen analisis bisa menyatakan emosi bersedih, senang, kecewa dan lain sebagainya[10].

* + 1. **Topic modeling**

Pemodelan topik merupakan salah satu teknik dalam text mining untuk penambangan data, penemuan data laten, dan mencari hubungan antar data atau dokumen teks. Tujuan topik modeling adalah untuk secara otomatis menentukan topik dari sekumpulan dokumen[6].

* + 1. **Twitter**

Twitter adalah salah satu sosial media yang sangat sering digunakan oleh masyarakat di Indonesia. Twitter umum digunakan oleh banyak kalangan masyarakat di Indonesia baik dari kalangan muda sampai kalangan tua karena praktis digunakan dan informasi yang diberikan cenderung cepat menyebar luas [10]. Informasi yang dapat dibagikan pada twitter maksimal berjumlah 280 karakter.

* + 1. **Scraping Data**

*Scraping data* adalah teknik untuk mengekstrak data dari World Wide Web atau biasa disingkat dengan WWW dan menyimpan ke file ataupun database sehingga dapat digunakan untuk pengolahan atau analiss nantinya [11]. Scraping data twitter digunakan untuk mengambil data tweet dalam jangka waktu data yang bisa ditentukan oleh penelitinya sendiri secara otomatis [2].

* + 1. **Preprocessing**

Preprocessing adalah teknik manipulasi data sebelum data diklasifikasikan. Dalam preprocessing data, data penting akan dipisahkan dari data yang tidak penting dan akan terjadi proses menghilangkan data mentah yang telah diperoleh sebelumnya [12]. Pada proses ini juga dilakukan *labeling data* menggunakan kamus sentimen berbahasa indonesia agar lebih akurat dan efektif dalam pengerjaan.

* + 1. **Lexicon-Based**

leksikon base adalah metode yang umum untuk menganalisis emosi di media sosial. Ini karena nyaman digunakan. Metode yang digunakan Lexicon Based adalah dengan menggunakan kamus sebagai bahasa atau sumber leksikon. Karena didasarkan sebagai pengklasifikasi emosi dari semua Opini, kumpulan emosi dapat dibagi menjadi kelas negatif dan positif[14]**.** Lexicon adalah kumpulan kata terdiri dari kata benda, kata keterngan, kata sifat, dan katakerja yang sering dijelaskan untuk delapan emosi (kegembiraan, kesedihan, kemarahan, ketakutan, jijik, kejutan, kepercayaan, dan antisipasi) serta untuk sentimen positif dan negative [15].

* + 1. **Vader**
    2. **Bag of Word**

Kantong kata atau Bag of Word (BoW) merupakan salah satu pendekatan yang terkenal cepat & gampang diterapkan, namun mengalami aliasing persepsi, terutama lantaran kuantisasi vektor[17].

* + 1. **Topic Modeling**
    2. **LDA**

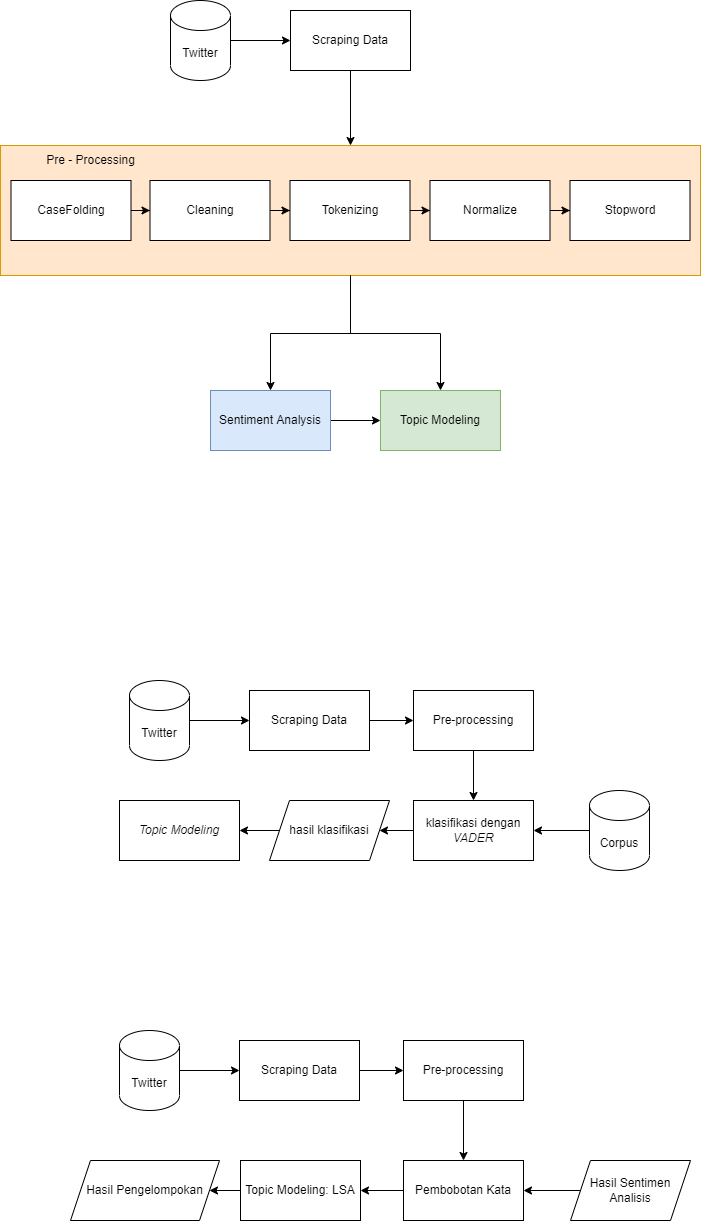
*Latent-Dirichlet Allocation* (LDA) adalah jenis algoritma yang paling umum dan paling banyak digunakan dan dipandang sebagai pendekatan yang efektif dalam proses pemodelan topik. LDA dapat menemukan struktur topik tersembunyi dengan cara menemukan campuran kata yang terkait dengan topik lain. Setiap dokumen memiliki beberapa topik. Jumlah topik yang akan dihasilkan dapat ditentukan sebelum proses dilakukan. Topik yang ditemukan dengan LDA memiliki tingkat konsistensi dan efektifan yang tinggi[18]. LDA adalah Salah satu metode pemodelan topik yang paling umum [19].

**BAB III**

**METODOLOGI PENELITIAN**

* 1. **Metode Penelitian**

Pada penyusunan tugas akhir, terdapat diagram alir sebagai berikut:



**Gambar 3.1 Flowchart Penelitian**

Berikut langkah-langkah penyelesaian penelitian ini yaitu:

* + 1. Pengumpulan Data

Pengumpulan data menggunakan teknik *Scraping data* dari salahsatu sosial media yaitu Twitter. Data yang dikumpulkan berupa Tweet dengan kata kunci “harga cabai naik” dan “harga cabe naik” dari Twitter berbahasa Indonesia.

* + 1. *Preprocesing*

Data yang telah diperoleh dari teknik *Scraping* perlu dilakukan Preprocessing. Proses ini akan mengolah data awal yang masih bermacam-macam bentuknya untuk dijadikan data teratur yang dapat diterapkan pada metode pengklasifikasian. Pada tahap ini dilakukan proses *Cleaning Data* dan juga *Case Folding.*

* + - 1. *Case Folding*

*Case folding* memiliki tujuan untuk merubah semuakata menjadi *lowercase*

* + - 1. *Cleaning*

*Cleaning* bertujuan untuk melakukan pembersihan data tagar, mention, tanda baca, angka, URL dan kata lainnya yang tidak berguna.

* + - 1. *Tokenizing*

*Tokenizing* digunakan untuk memecahan kalimat menjadi bagian-bagian atau token.

* + - 1. *Stopwods Removing*

*Stopword Removal* digunakan untuk membuang kata yang tidak diperlukan berdasarkan kamus *Stopword* Bahasa Indonesia.

* + - 1. *Normalizatioan*

*Normalization* bertujuan untuk menormalisasikan kata yang tida baku menjadi katabaku berdasarkan kamus Normalisasi Bahasa Indonesia.

* + 1. Pelebelan sentiment

Proses ini bertujuan untuk memberikan label pada opini berupa positif atau negatif secara otomatis menggunakan *Vader*.

* + 1. Pembobotan Kata

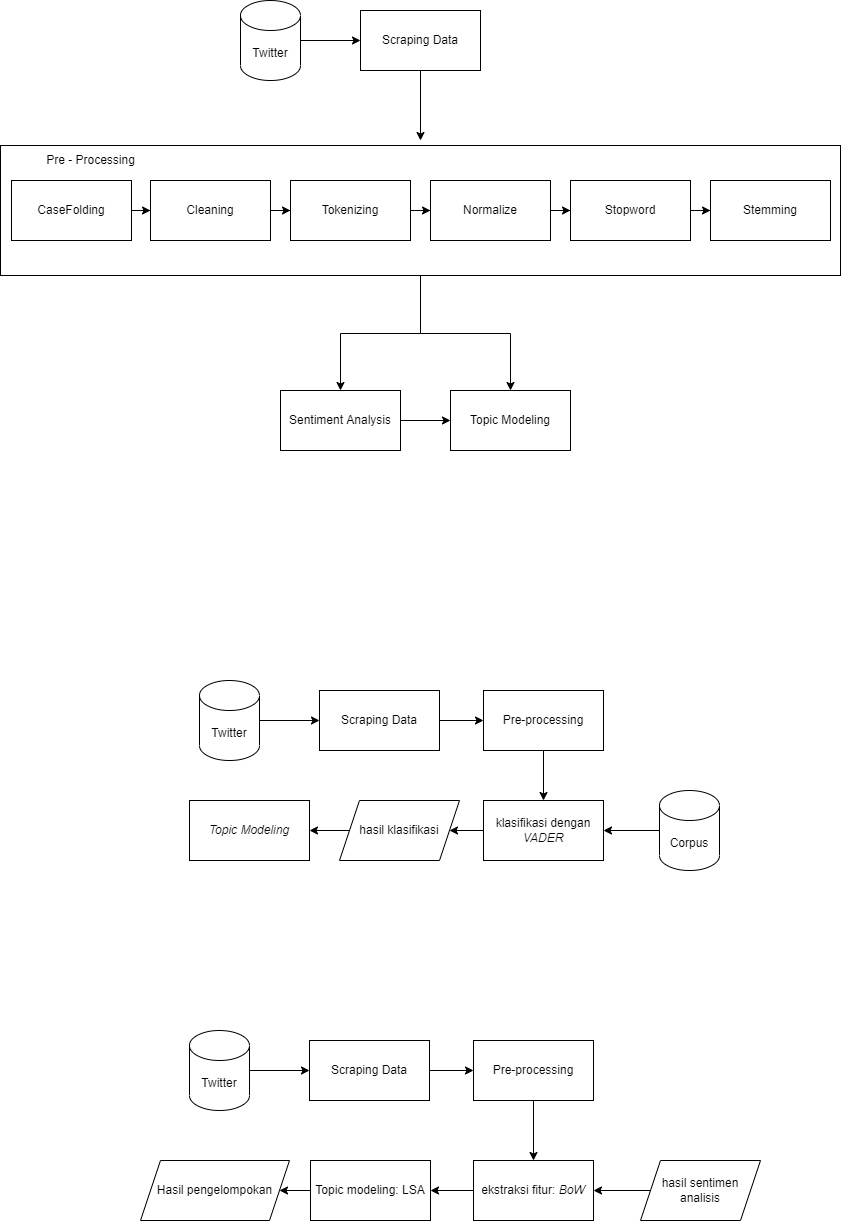
Setelah mendapatkan kata yang sudah teratur dengan metode sebelumnya, selanjutnya pembobotan kata menggunakan metode BoW atau *Bag of Word* dimana metode ini berguna sebagai menghitung nilai/bobot setiap kata.

* + 1. Mengelompokan Data

Pada proses ini akan dilakukan pengelompokan data dengan

* 1. **Metode Analisis Sentimen**

Berikut ini merupakan alur dari proses Analisis Sentimen

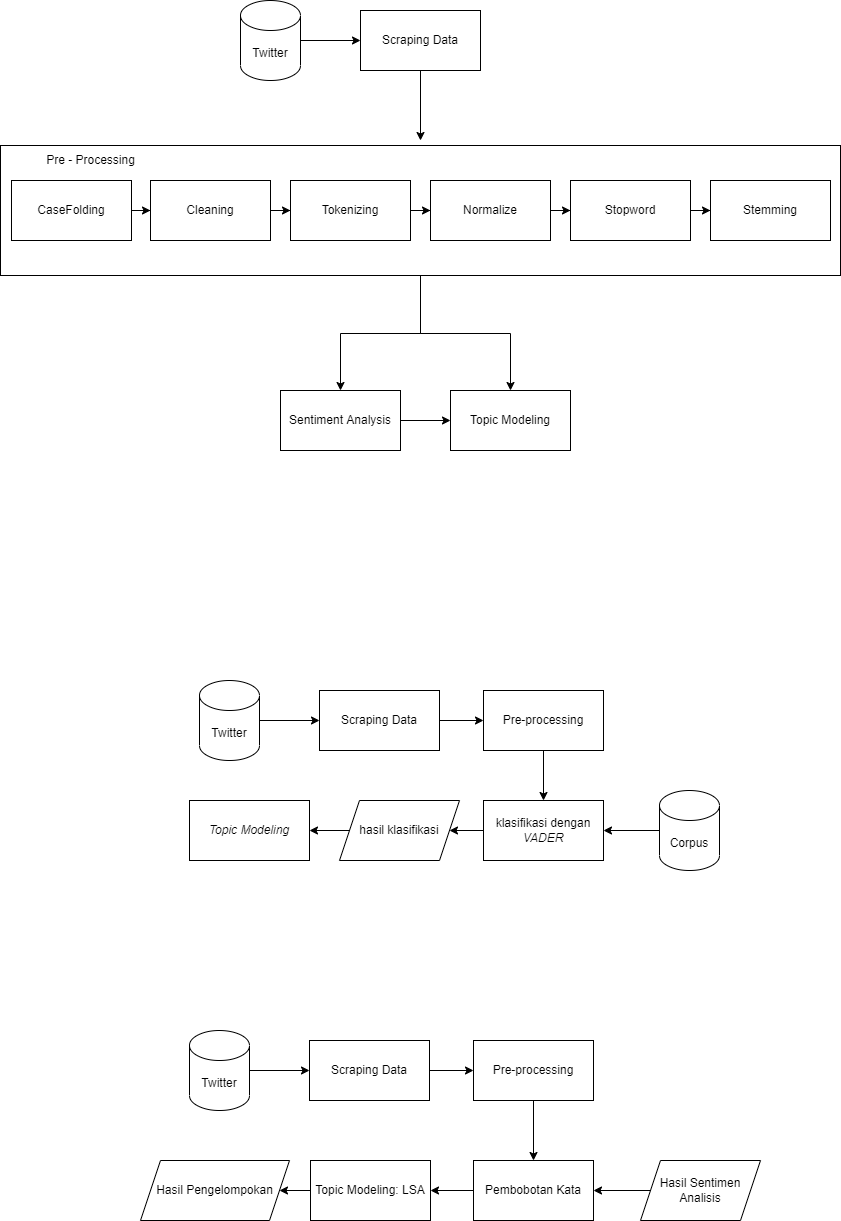
****

**Gambar 3.2 Flowchart Analisis Sentimen**

Gambar 3.2 meenunjukan alur proses sentiment analisis pada penelitian ini. Dimulai dengan scraping data menggunakan library yang terdapat pada python, lalu dilanjutkan ke tahap *preproscessing* data menggunakan *Case Folding, Cleansing, Tokenizing, Stopword Removal, Normalization* dan *Stemming,* kemudian dilanjutkan dengan klasifikasi *Lexicon-Base* menggunakan *Vader* yang data lexiconnya diambil dari *corpus*. Hasil dari klasifikasi *Lexicon-Base* berupa sentiment positif dan sentimen negative. Hasil klasifikasi kemudian dilanjutkan ke tahapan pemodelan topik untuk dilihat topiknya per sentiment.

* 1. **Metode Topic Modeling**

Berikut ini merupakan alur dari proses Topic Modeling



**Gambar 3.3 Flowchart Topic Modeling**

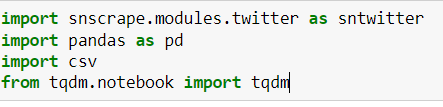
Pada gambar 3.3 menunjukan proses pemodelan topik. Proses ini kurang lebih sama seperti proses sentiment analisis, dimulai dari *scraping* data menggunakan library yang terdapat pada python, lalu dilanjutkan ke tahap *preproscessing* data, kemudian dilanjutkan tahap ekstraksi fitur menggunakan *Bag of Word,* pada tahap ini juga dilakukan secara parallel ekstraksi fitur dari data hasil sentiment analisis. Setelah melakukan ektraksi fitur, dilanjutkan ke tahap Pemodelan Topik menggunakan LDA, metode ini dipilih karna merupakan salah satu metode yang sering digunakan.

**BAB IV**

**PEMBAHASAN**

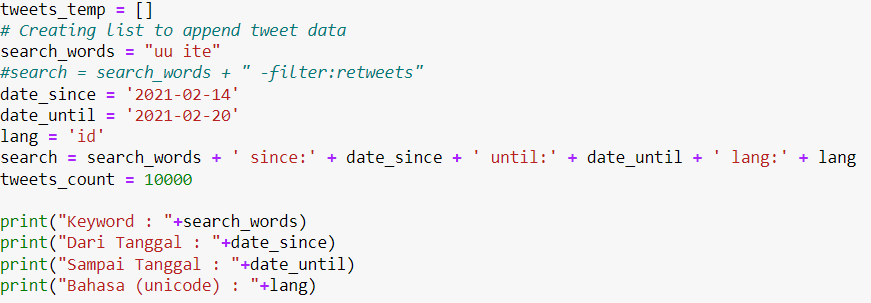
* 1. ***Scraping data***

Pada tahapan ini dilakukan pengumpulan data menggunakan Teknik *scraping.* Data twitter dikumpulkan menggunakan *Library snscrape*. Snscrape adalah *Library* yang berisi berbagai fungsi untuk mengumpulkan tweet dari Twitter. Snscrape berguna untuk membantu memfilter tweet pada kondisi dibutuhkan seperti berdasarkan *username*, tanggal, bahasa dll[20]. Library yang digunakan pada tahapan ini sebagai berikut:

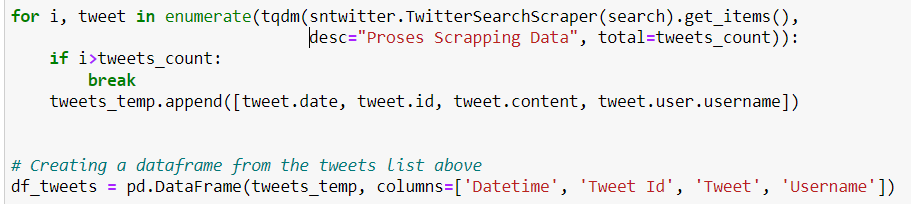


**Gambar 4.1 Library Tahap Scrapping.**

Keemapat *library* tersebut digunakan untuk *scraping* data pada sosial media twitter. Setelah itu, memasukan data yang digunakan dalam penelitian ini misalnya untuk tanggal pengambilan data dimulai dari tanggal 14 sampai 20 februari 2021 dengan kata kunci “UU ITE” dengan pengaturan Bahasa Indonesia seperti gambar berikut:



**Gambar 4.2 Syntax Pencarian Tweet**



**Gambar 4.3 Proses Mendapatkan Data**

Setelah data terkumpul, dilakukan filtrasi data yang duplikat. Tahapan ini dilakukan untuk menghilangkan data yang berulang terhimpun, cara menjalankannya seperti berikut:

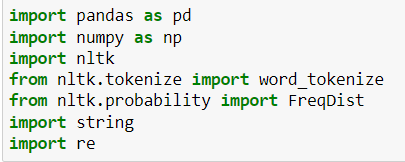


**Gambar 4.4 Drop Data Duplikat**

Pada penarikan ini terhimpun 10.001 data tweet dari tanggal 14 sampai 20 februari 2021. Tanggal tersebut diambil karna tanggal tersebut merupakan pencarian tertinggi mengenai UU ITE berdasarkan *Google Trends.*

* 1. ***Pre-processing***

Pada tahapan ini dilakukan pembersihan data yang telah dikumpulkan pada tahap sebelumnya. Tahapan ini bertujuan untuk membersihkan data agar dapat diproses pada untuk tahap selanjutnya. Berikut merupakan *Library* yang digunakan pada tahap ini:



**Gambar 4.5 Library Tahap Pre-processing**

* + 1. ***Case Folding***

Tahapan ini berujuan untuk mengubah semua data menjadi huruf kecil seperti Gambar 4.6 dan hasilnya pada Tabel 4.1 berikut ini.



**Gambar 4.6 Proses *Case Folding***

|  |  |
| --- | --- |
| Sebelum | Sesudah |
| Demi menyelamatkan seorang abu janda agar tidak terjerah hukum, negara harus mengalah dengan merevisi UU ITE. | demi menyelamatkan seorang abu janda agar tidak terjerah hukum, negara harus mengalah dengan merevisi UU ITE. |

**Tabel 4.1 Hasil *Case Folding***

* + 1. ***Cleaning***

Selanjutnya pada tahapan ini dilakukan cleaning untuk menghilangkan data tagar, mention, tanda baca, angka dan URL yang tidak berguna seperti Gambar 4.7 dan hasilnya pada **Tabel 4.2** berikut ini:



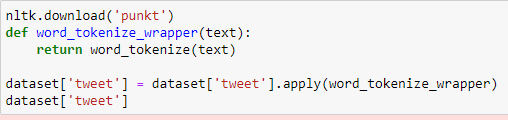
**Gambar 4.7 Proses *Cleaning***

|  |  |
| --- | --- |
| Sebelum | Sesudah |
| dibeberkan roy suryo, ternyata ini sejarah lahirnya UU ITE, ada 2 poin penting https://t.co/l4tozlwmt4 | dibeberkan roy suryo ternyata ini sejarah lahirnya UU ITE ada poin penting |
| "@keuangannews\_id @budimandjatmiko benar enggak mas? saya nggak percaya, kelasnya mas budi nggak mungkin mundur. pasti bang @fahrihamzah jg gx bakal percaya kalau mas bud nggak berani ditantang debat UU ITE" | benar enggak mas saya nggak percaya kelasnya mas budi nggak mungkin mundur pasti bang jg gx bakal percaya kalau mas bud nggak berani ditantang debat UU ITE |

**Tabel 4.2 Hasil *Cleaning***

* + 1. ***Tokenizing***

Kemudian dilanjutkan ketahap tokenizing yang bertujuan untuk memecahan kalimat menjadi bagian-bagian atau token. Hal ini berfungsi untuk menjalankan tahapan lainnya seperti normalize, stopword removal dll. Berikut syntax yang digunakan seperti Gambar 4.8 dan hasilnya pada Tabel 4.2 berikut ini:



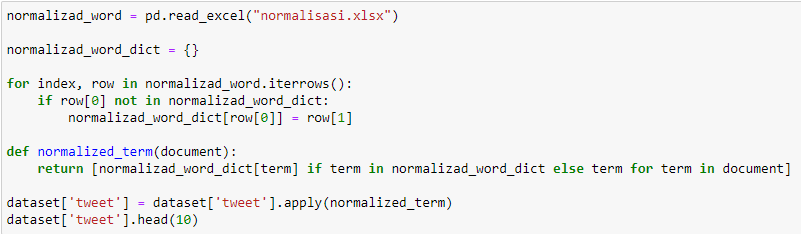
**Gambar 4.8 Proses *Tokenizing***

|  |  |
| --- | --- |
| Sebelum | Sesudah |
| dibeberkan roy suryo ternyata ini sejarah lahirnya UU ITE ada poin penting | ['dibeberkan', 'roy', 'suryo', 'ternyata', 'ini', 'sejarah', 'lahirnya', 'uu', 'ite', 'ada', 'poin', 'penting'] |

**Tabel 4.2 Hasil *Tokenizing***

* + 1. ***Normalize***

Setelah melakukan tokenizing, dilanjutkan ke tahap normalize yang berguna untuk mengubah kata tidak baku menjadi kata baku seperti Gambar 4.9 dan hasilnya pada Tabel 4.3 berikut ini:



**Gambar 4.9 Proses Normalisasi**

|  |  |
| --- | --- |
| Sebelum | Sesudah |
| "['ga', 'malu', 'koar', 'bgt', 'lupa', 'ya', 'klo', 'uu', 'ite', 'merupakan', 'salah', 'satu', 'produk', 'pks', 'ketika', 'berkuasa', 'pks', 'beri', 'waktu', 'sebulan', 'agar', 'jokowi', 'revisi', 'uu', 'ite', 'muannas', 'alaidid', 'mestinya', 'mardani', 'ali', 'sera', 'malu', 'seputar', 'tangsel']" | "['ga', 'malu', 'teriak', 'banget', 'lupa', 'iya', 'kalau', 'uu', 'ite', 'merupakan', 'salah', 'satu', 'produk', 'pks', 'ketika', 'berkuasa', 'pks', 'beri', 'waktu', 'sebulan', 'agar', 'jokowi', 'revisi', 'uu', 'ite', 'muannas', 'alaidid', 'mestinya', 'mardani', 'ali', 'sera', 'malu', 'seputar', 'tangsel']" |

**Tabel 4.3 Hasil Normalisasi**

* + 1. ***Stopword Removal***

Tahapan ini memiliki fungsi untuk membuang kata yang tidak diperlukan dalam sentiment analisis. Berikut syntax yang digunakan seperti Gambar 4.10 dan hasilnya pada Tabel 4.4 berikut ini:



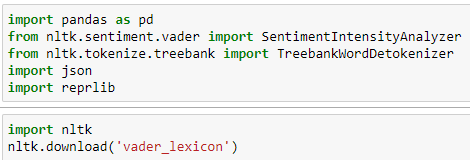
**Gambar 4.10 Proses** ***Stopword Removal***

|  |  |
| --- | --- |
| Sebelum | Sesudah |
| "['ga', 'malu', 'teriak', 'banget', 'lupa', 'iya', 'kalau', 'uu', 'ite', 'merupakan', 'salah', 'satu', 'produk', 'pks', 'ketika', 'berkuasa', 'pks', 'beri', 'waktu', 'sebulan', 'agar', 'jokowi', 'revisi', 'uu', 'ite', 'muannas', 'alaidid', 'mestinya', 'mardani', 'ali', 'sera', 'malu', 'seputar', 'tangsel']" | "['malu', 'teriak', 'banget', 'lupa', 'uu', 'ite', 'salah', 'produk', 'pks', 'berkuasa', 'pks', 'sebulan', 'jokowi', 'revisi', 'uu', 'ite', 'muannas', 'alaidid', 'mestinya', 'mardani', 'ali', 'sera', 'malu', 'seputar', 'tangsel']" |

**Tabel 4.4 Hasil *Stopword Removal***

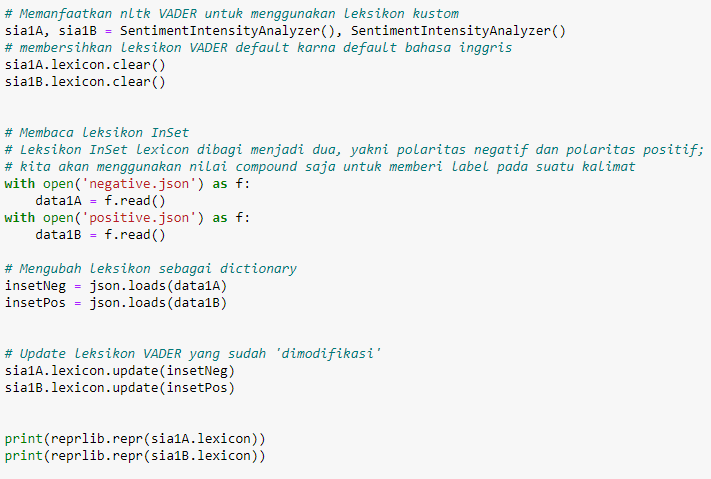
* 1. ***Sentiment Analysis***

Setelah tahapan *Pre-processing* selesai dilakukan, dilanjutkan ketahap Analisis sentimen. Tahapan ini dilakukan untuk mengklasifikasi sentimen dari data yang didapat. Proses klasifikasi yang digunakan adalah klasifikasi jenis Lexicon-based dengan Library yang digunakan seperti berikut:



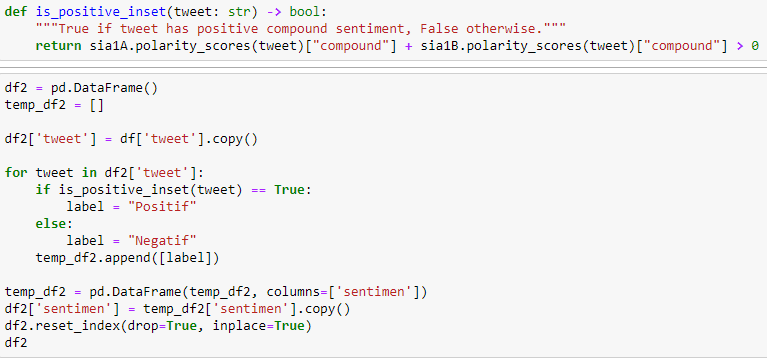
**Gambar 4.12 Library Tahap Analisis Sentimen**

Pada proses ini menggunakan kamus Bahasa Indonesia yang telah dibuat pada penelitian *Inset Lexicon.* Kamus ini berisi kata-kata Bahasa Indonesia yang sudah diberi nilai dari -5 sampai +5 untuk setiap kata. Berikut syntax yang digunakan seperti Gambar 4.13:

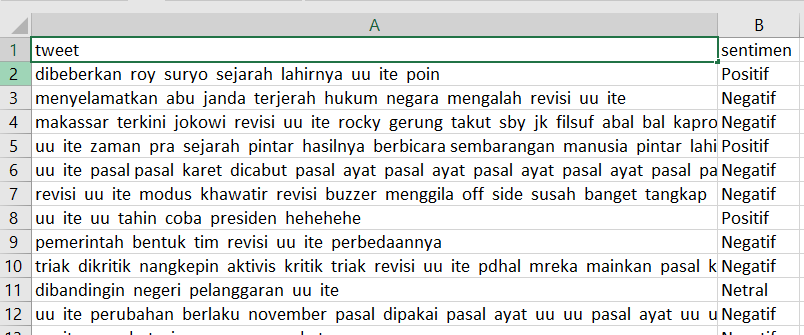


**Gambar 4.13 Memasukan Lexicon Inset kedalam Variable**

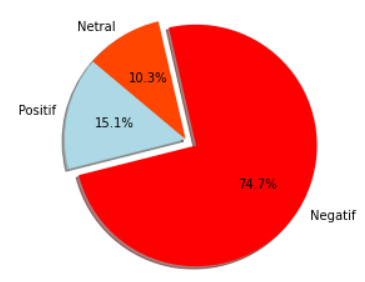
Proses ini menggunakan *Vader* untuk melakukan pelabelan secara otomatis dengan menggunakan data dari kamus InSet Lexicon yang sudah dimasukan sebelumnya seperti Gambar 4.14 dan hasilnya seperti Gambar 4.15



**Gambar 4.14 Prosen Pelebelan**



**Gambar 4.15 Hasil Pelebelan**



**Gambar 4.16 Hasil Analisis Sentimen**

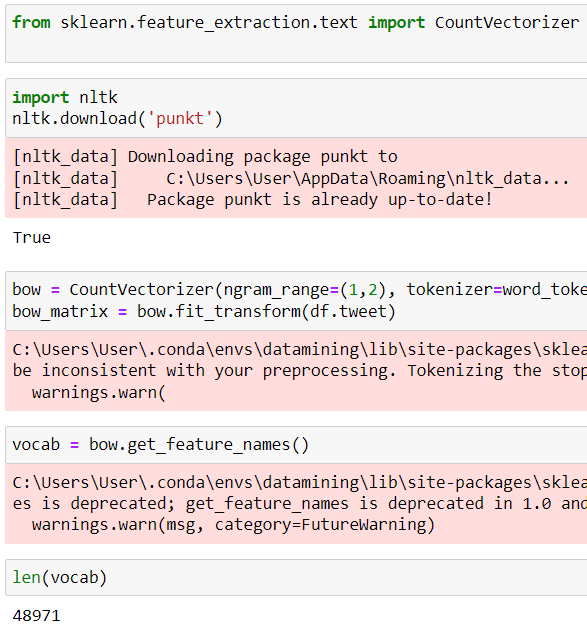


**Gambar 4.17 Visualisasi Wordcloud**

Dapat dilihat pada Gambar 4.16 bahwa sentiment masyarakat mengenai UU ITE adalah negative sebesar 74,7%. Kemudian, pada Gambar 4.17 terlihat bahwa kata yang paling banyak muncul adalah “uu”, “ite” dan “revisi”. Kemudian, terdapat beberapa kata *collocation* yang muncul, yaitu “UU ITE”, “revisi uu” & “pasal karet”.

* 1. **Pembobotan Kata**

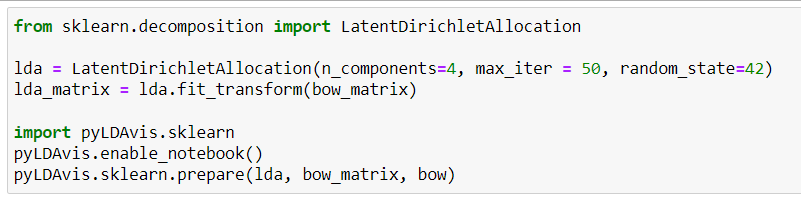
Pada proses ini dilakukan untuk merubah kata-kata yang telah dihimpun menjadi vector agar dapat diproses pada proses berikutnya. Proses ini menggunakan *Bag of Word* karna cocok dengan Teknik visualisasi yang akan digunakan pada penelitian ini. Berikut syntax yang digunakan seperti Gambar 4.18:



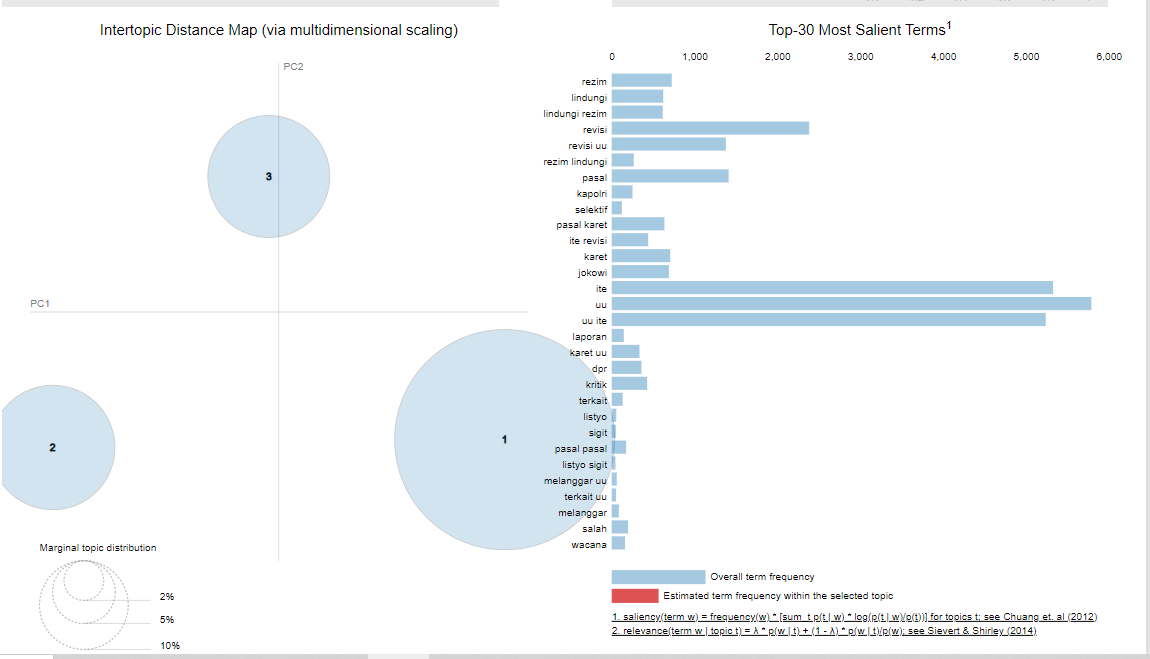
**Gambar 4.18 Proses Tahap Ekstraksi Fitur**

* 1. ***Topic Modeling***

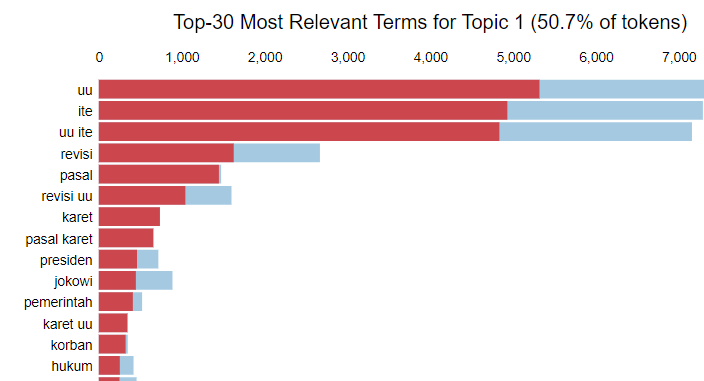
Setelah dilakukan tahap ekstraksi fitur, data siap untuk masuk ketahap pemodelin topik. Tahapan ini memiliki tujuan untuk menemukan topik yang sedang dibicarakan. Pada penelitian ini menggunakan metode LDA untuk proses pengelompokan topik dan library Pyldavis untuk visualisasinya seperti Gambar 4.19 dan hasilnya pada Gambar 4.20, Gambar 4.21 dan Gambar 4.22 berikut ini:



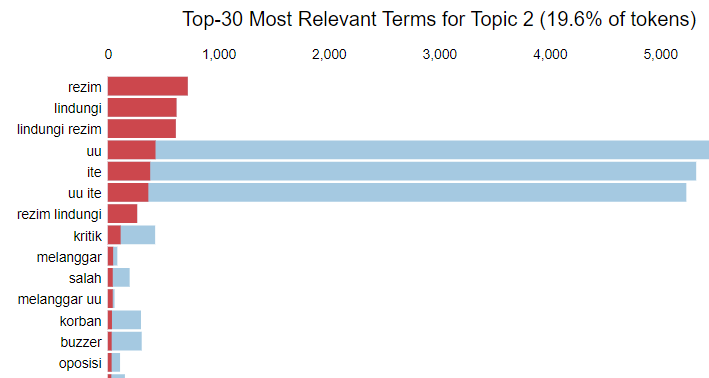
**Gambar 4.19 Proses Tahap Topic Modeling**



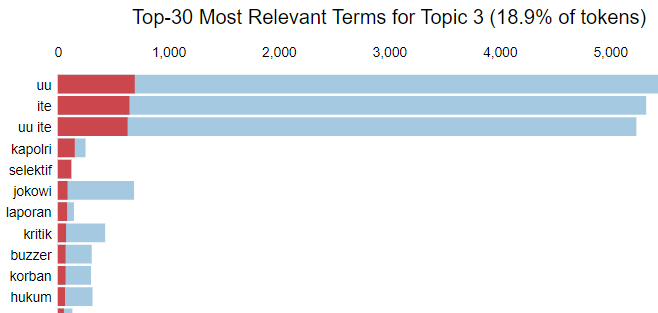
**Gambar 4.20** **Visualisasi Topic Modeling Negatif**



**Gambar 4.21 Topik 1**

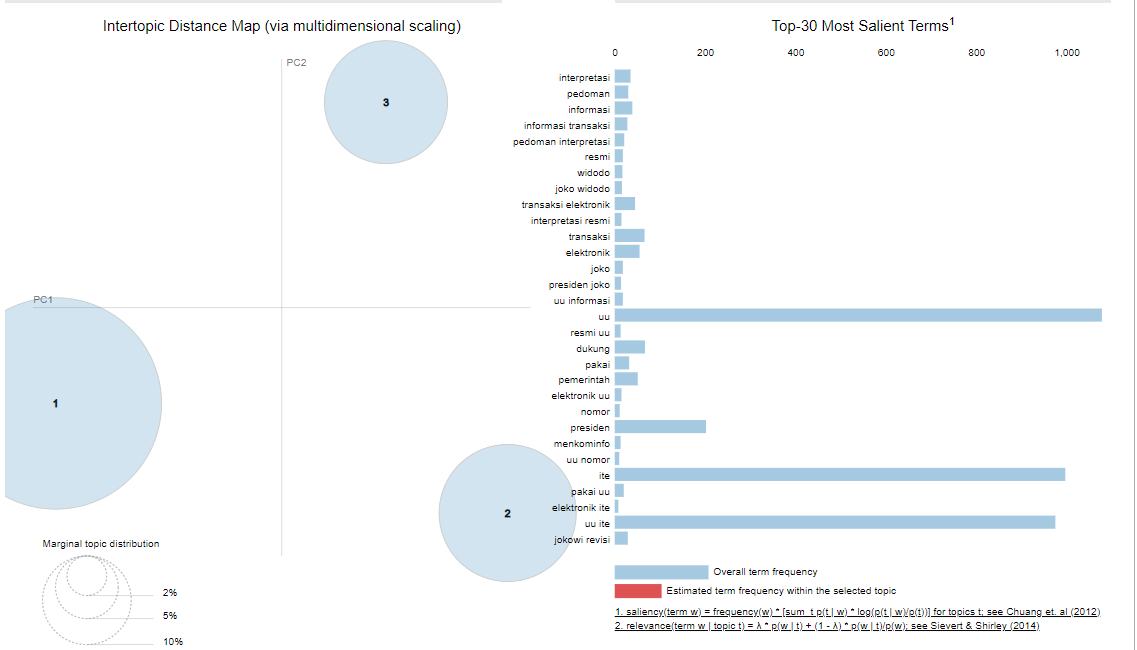


**Gambar 4.22 Topik 2**

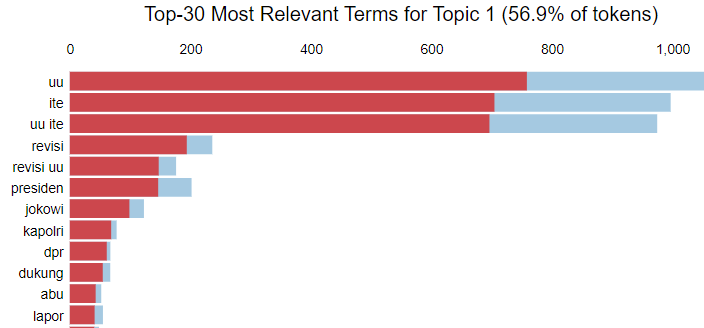


**Gambar 4.23 Topik 3**

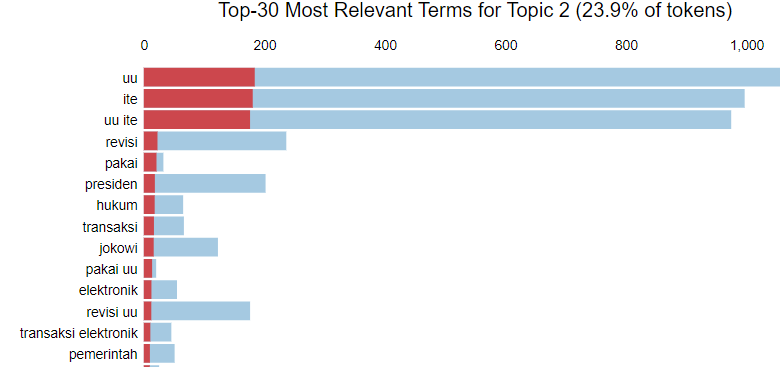
Pada Gambar 4.20 dapat dilihat terdapat 3 topik mengenai UU ITE dengan sentiment negative. Pada topik 1 membicarakan topik mengenai



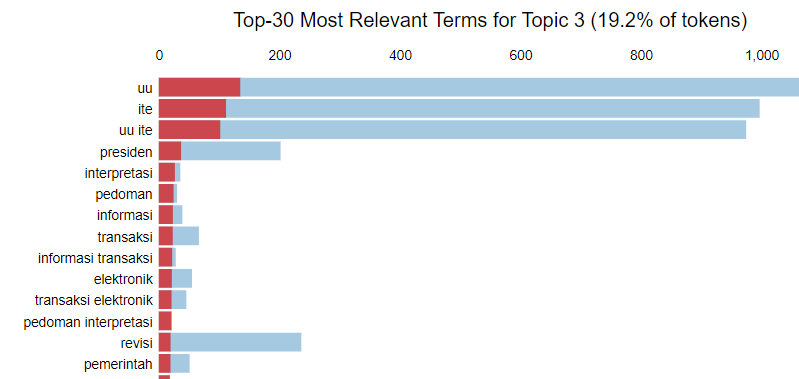
**Gambar 4.24 Visualisasi Topic Modeling Positif**



**Gambar 4.25 Topik 1**

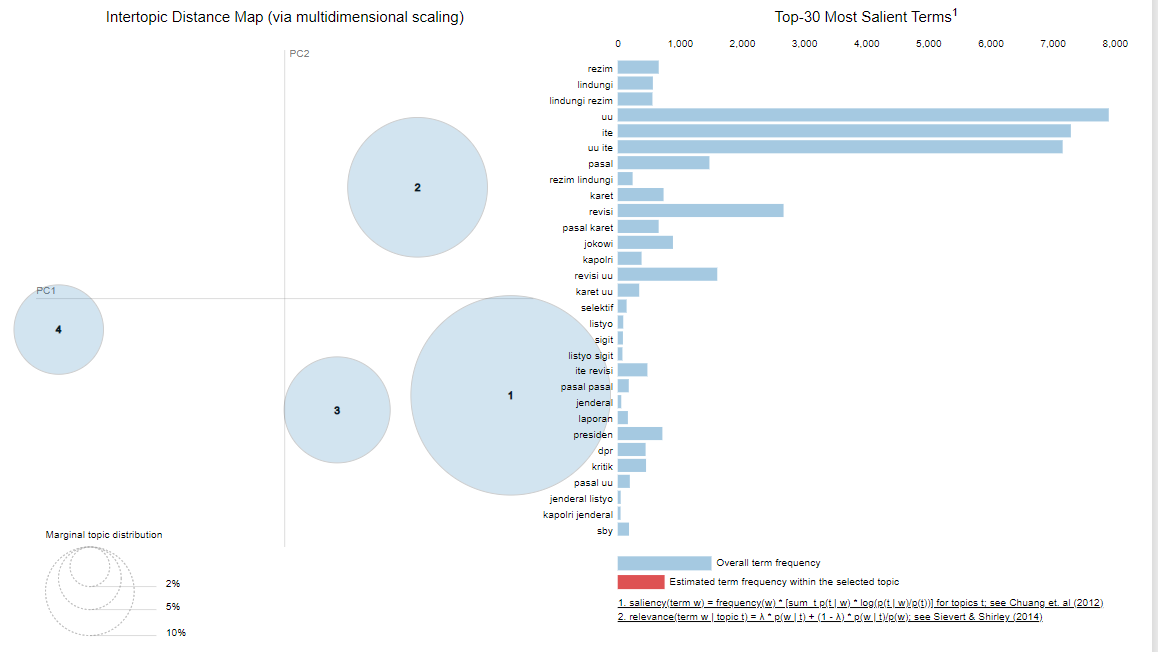


**Gambar 4.26 Topik 2**

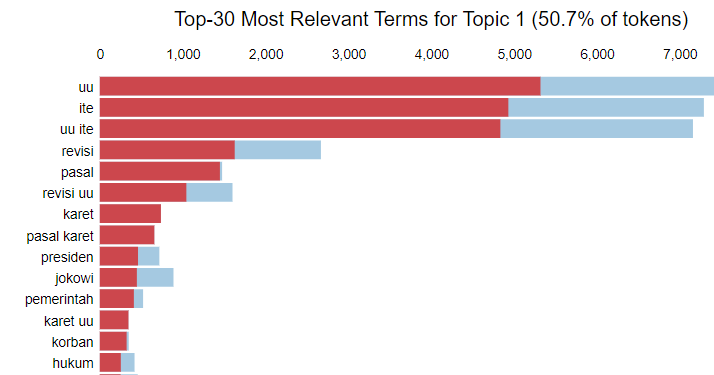


**Gambar 4.27 Topik 3**

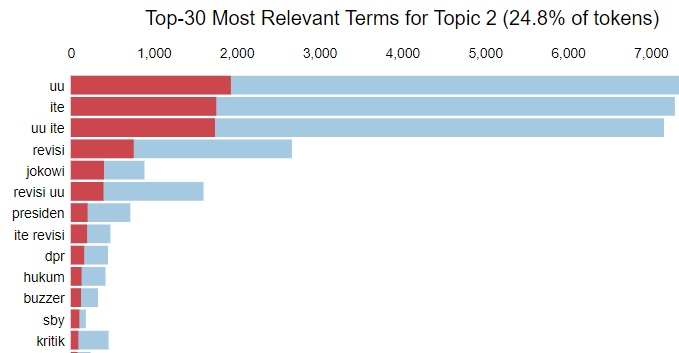
Lalu pada Gambar 4.24 juga terbagi menjadi 3 topik besar mengenai UU ITE dengan sentiment positif. Pada topik pertama data membahas mengnai permintaan revisi UU ITE kepada pemangku jabatan, terlihat ada beberapa pemangku jabatan yang disebutkan dalam topik tersebut seperti presiden, kapolri dan DPR. Lalu pada topik 2 membahas mengenai permintaan revisi UU ITE kepada presiden mengenai transaksi elektronik. Kemudian pada topik 3 membahas mengenai setujunya masyarakat untuk dibuatnya interpretasi terhadap UU ITE.



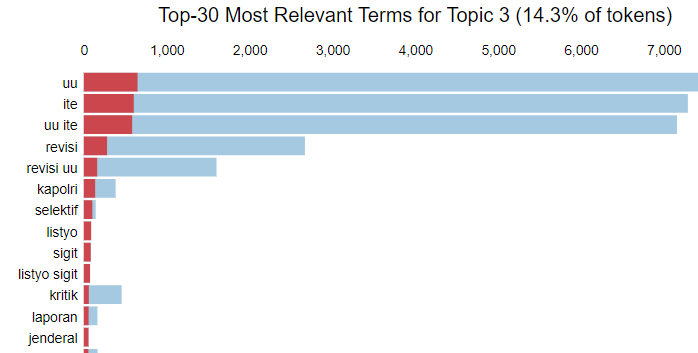
**Gambar 4.28 Visualisasi Topic Modeling Keseluruhan**



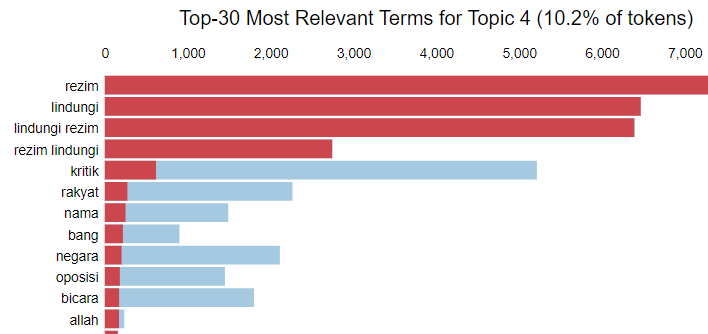
**Gambar 4.29 Topik 1**



**Gambar 4.30 Topik 2**



**Gambar 4.31 Topik 3**



**Gambar 4.32 Topik 4**

Kemudian pada Gambar 4.22 yang dibicarakan secara garis besar adalah mengenai topik revisi pasal UU ITE mengenai pasal kritik dan pembuatan pedoman interpretasi UU ITE.

**BAB V**

**KESIMPULAN DAN SARAN**

* 1. **Kesimpulan**

Dari data yang terhimpun sebanyak 10001 data, tersisa 8765 data tweet yang siap untuk digunakan dalam penelitian ini. Hasil sentiment dari menelitian ini adalah negative sebesar 74,6% setelah dilakukan klasifikasi menggunakan *Vader.* Hal ini dapat mengindikasikan bahwa adanya ketidakpuasan dari masyarakat mengenai kebijakan tersebut. Untuk visualisasi dapat dilihat topik yang paling sering dibicarakan yaitu “uu” dan “ite”, disusul dengan “revisi”, “pasal” dan “karet”

Pada proses topik modeling, ditemukan bahwa iterasi terbaik pada penelitian ini adalah sebanyak 50 kali sehingga didapatkan topik yang maksimal. Kesimpulan dari topik yang dibahas pada penelitian ini adalah masyarakat meminta kepada pemangku jabatan untuk merevisi Kembali kebijakan UU ITE karna dirasa bersifat karet dan anti kritik serta sudah banyak memakan korban. Karna permasalahan ini, presiden dan kapolri berencara akan merevisi dan membuat pedoman interpretasi UU ITE

* 1. **Saran**

Penelitian ini hanya menggunakan 1 periode penarikan data karna sampai saat penelitian ini dibuat, UU ITE masih belum direvisi. Pada penelitian selanjutnya diharapkan untuk melakukan 2 periode penarikan untuk membandingkan sentiment masyarakat mengenai UU ITE.

**Daftar Referensi**

[1] A. P. Thenata, “Text Mining Literature Review on Indonesian Social Media,” vol. 7, no. 2, pp. 226–232, 2021.

[2] I. Zulfa and E. Winarko, “Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017.

[3] J. Raya, C. Cisaat, C. Kaler, S. Regency, and J. Barat, “Analisis Sentimen Seputar UU ITE Menggunakan Algoritma Support Vector.”

[4] M. G. Undap, V. P. Rantung, and P. T. D. Rompas, “Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based,” 2021.

[5] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, “Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm,” *J. Phys. Conf. Ser.*, vol. 1821, no. 1, 2021, doi: 10.1088/1742-6596/1821/1/012054.

[6] K. D. A. Salsabila and N. Trianasari, “Analisis Persepsi Produk Kosmetik Menggunakan Metode Sentiment Analysis dan Topic Modeling (Studi Kasus: Laneige Water Sleeping Mask),” *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 1–9, 2021, doi: 10.26905/jtmi.v7i1.5593.

[7] C. S. Hudaya, H. Fakhrurroja, and A. Alamsyah, “Analisis Persepsi Konsumen Terhadap Brand Go-Jek Pada Media Sosial Twitter Menggunakan Metode Sentiment Analysis Dan Topic Modelling,” *J. Mitra Manaj.*, vol. 3, no. 6, pp. 664–673, 2019, doi: 10.52160/ejmm.v3i6.244.

[8] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, “Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.

[9] I. K. Susanto, “Analisis Sentimen dan Topic Modelling Pada Pembelajaran Online di Indonesia Melalui Twitter,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 6, no. 2, p. 85, 2021, doi: 10.31328/jointecs.v6i2.2350.

[10] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017.

[11] B. Zhao, “Encyclopedia of Big Data,” *Encycl. Big Data*, no. May 2017, 2020, doi: 10.1007/978-3-319-32001-4.

[12] J. Jtik, J. Teknologi, R. T. Aldisa, and M. A. Abdullah, “Analisis Sentimen Mengenai Vaksin Sinovac di Media Sosial Twitter Menggunakan Metode Naïve bayes Classification,” vol. 6, no. 3, pp. 1–5, 2022.

[13] A. F. Hidayatullah, C. I. Ratnasari, and S. Wisnugroho, “Analysis of Stemming Influence on Indonesian Tweet Classification,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 14, no. 2, pp. 665–673, 2016, doi: 10.12928/TELKOMNIKA.v14i1.3113.

[14] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, “ANALISIS SENTIMEN PENGGUNA GOPAY MENGGUNAKAN METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE,” 2019, [Online]. Available: http://studentjournal.umpo.ac.id/index.php/komputek.

[15] S. Kiritchenko, X. Zhu, and S. M. Mohammad, “Sentiment analysis of short informal texts,” *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, 2014, doi: 10.1613/jair.4272.

[16] Y. D. Prabowo, T. L. Marselino, and M. Suryawiguna, “Pembentukan Vector Space Model Bahasa Indonesia Menggunakan Metode Word to Vector,” *J. Buana Inform.*, vol. 10, no. 1, p. 29, 2019, doi: 10.24002/jbi.v10i1.2053.

[17] N. Kejriwal, S. Kumar, and T. Shibata, “High performance loop closure detection using bag of word pairs,” *Rob. Auton. Syst.*, vol. 77, pp. 55–65, 2016, doi: 10.1016/j.robot.2015.12.003.

[18] C. Zou, “Analyzing research trends on drug safety using topic modeling,” *Expert Opin. Drug Saf.*, vol. 17, no. 6, pp. 629–636, 2018, doi: 10.1080/14740338.2018.1458838.

[19] P. Damayanti, D. Purwitasari, and N. Suciati, “Eliminasi Non-Topic Menggunakan Pemodelan Topik untuk Peringkasan Otomatis Data Tweet dengan Konteks Covid-19,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 199, 2021, doi: 10.25126/jtiik.0814324.

[20] T. Sarkar and N. Rajadhyaksha, “TLA: Twitter Linguistic Analysis,” pp. 1–7, 2021, [Online]. Available: http://arxiv.org/abs/2107.09710.