

Analisis Sentimen Opini Publik dalam Bahasa Indonesia Terhadap UU Cipta Kerja Menggunakan Naïve Bayes

Alvin Andreas¹⁾, Yulius Denny Prabowo²⁾

Informatika, Fakultas Ilmu Komputer dan Desain, Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210

¹⁾ Email: 2017103369@student.kalbis.ac.id

²⁾ Email: yulius.prabowo@kalbis.ac.id

Abstract: This study aims to apply the Naïve Bayes method in analyzing public review of the Job Work and measuring the accuracy of Naïve Bayes using twitter dataset. The dataset used in this study was taken from the data text on Twitter social media with 160 tweets. The dataset will be classified into positive sentiments and negative sentiments. The method used in the classification process is the Naïve Bayes method. The software development method used in this research is the Incremental method. The accuracy results obtained during the study were 85% with an average precision, recall, and f1-score of 85%.

Keywords: Incremental, Naïve Bayes, Sentiment Analysis, Twitter

Abstrak: Penelitian ini bertujuan untuk menerapkan metode Naïve Bayes dalam menganalisis ulasan masyarakat terhadap UU Cipta kerja dan mengukur akurasi Naïve Bayes dengan menggunakan dataset twitter. Dataset yang digunakan dalam penelitian ini diambil dari data teks pada media sosial Twitter dengan total 160 cuitan. Dataset tersebut akan diklasifikasikan menjadi sentimen positif dan negatif. Metode yang digunakan dalam proses klasifikasi adalah metode Naïve Bayes. Metode pengembangan perangkat lunak yang digunakan dalam penelitian ini adalah metode Incremental. Hasil akurasi yang didapatkan saat penelitian berlangsung sebesar 85% dengan rata - rata precision, recall, dan f1-score sebesar 85%.

Kata kunci: Incremental, Naïve Bayes, Sentiment Analysis, Twitter

I. PENDAHULUAN

Omnibus law pertama kali disebutkan dalam pidato perdana Presiden Republik Indonesia periode ke-2 pada hari Minggu, 20 Oktober 2019. Menurut pemerintah, metode *omnibus law* ini dilakukan untuk menyusun kembali atau merevisi beberapa undang – undang bahkan puluhan yang salah satunya seperti sektor ekonomi yang diharapkan dapat meningkatkan posisi Indonesia terkait kemudahan membuka atau membangun sebuah usaha dan membangun Indonesia supaya menjadi salah satu dari kekuatan ekonomi dunia, mengurangi tingkat kemiskinan, dan mempunyai tenaga kerja yang memiliki

kualitas [1]. Pemerintah menganggap beberapa undang – undang yang sudah ada menghambat hal tersebut terjadi. Secara keseluruhan *omnibus law* yang diajukan oleh pemerintahan Indonesia terkait undang – undang Cipta Kerja dan Perpajakan. Omnibus Law RUU Cipta Kerja ini mengandung 11 klaster yang terdiri dari 15 bab dan 174 pasal.

Opini publik terkait *Omnibus Law* UU Cipta Kerja menjadi perdebatan di kalangan masyarakat karena terjadi pro dan kontra. Pro dan kontra yang terjadi di kalangan masyarakat karena metode *omnibus law* ini adalah sebuah hal yang baru bagi Negara Indonesia. Salah satu media sosial yang menjadi wadah masyarakat Indonesia untuk

mengeluarkan opini atau pendapatnya terkait permasalahan tersebut adalah Twitter. Menurut Berita Kompas.com kata kunci dengan tagar BatalkanOmnibusLaw dan kata kunci lain yang berhubungan dengan *omnibus law* menjadi *trending topic* di Twitter sejak Senin, 5 Oktober 2020. *Trending topic* terjadi setelah beberapa lembaga pemerintah menyetujui RUU *Omnibus Law* Cipta Kerja saat rapat paripurna DPR pada Januari 2020.

Beberapa golongan masyarakat menilai bahwasannya RUU tersebut terlalu dibuat terburu – buru dan merugikan beberapa masyarakat seperti buruh, pegawai kontrak, dan lain – lain. Sepanjang pengesahan *Omnibus Law* UU Cipta Kerja sampai setelah disahkan terjadi beberapa kali demo yang dilakukan oleh beberapa kelompok seperti mahasiswa, buruh, dan masih banyak lagi dengan skala demo yang cukup besar yang menuntut pemerintah untuk melakukan analisis kembali terkait isi dari *omnibus law*, menolak *omnibus law* dan menginginkan Presiden untuk mengeluarkan Perppu. Beberapa kalangan masyarakat juga menyambut baik *omnibus law* seperti yang dituliskan dalam cuitan salah satu pengguna Twitter DrunTamvan yang menyebut UU Cipta Kerja menjembatani kemudahan investasi di Indonesia.

Peneliti melihat permasalahan tersebut dan tertarik untuk mengetahui jumlah pro dan kontra dari sentimen yang dikeluarkan oleh masyarakat. Peneliti ingin melakukan pengelompokan sentimen – sentimen yang dikeluarkan masyarakat terhadap UU Cipta kerja namun karena banyaknya cuitan yang dikeluarkan masyarakat terakit topik ini akan membutuhkan waktu yang banyak dalam melakukan analisis sentimen secara manual sehingga peneliti akan membuat aplikasi yang efisien untuk dapat melakukan analisis sentimen

secara otomatis yang dilakukan oleh mesin. Metode klasifikasi yang akan digunakan dalam penelitian ini adalah Naïve Bayes dan peneliti tertarik untuk melihat akurasi yang didapatkan dari proses analisis sentimen menggunakan dataset Twitter menggunakan metode Naïve Bayes. Media sosial yang akan digunakan peneliti untuk mengumpulkan dataset dalam penelitian ini adalah Twitter.

Analisis sentimen sebelumnya sudah pernah dilakukan oleh Yassine AL Amrani, Mohamed Lazaar, Kamal Eddine EL Kadiri dengan menggunakan klasifikasi Support Vector Machine dan Random Forest dengan *recall*, *precision*, dan *f-measure* secara berurutan sebesar 83.4%, 83,4%, dan 83.4% [2]. Penelitian lainnya dilakukan oleh Soudamini Hota dan Sudhir Pathak dengan menggunakan klasifikasi KNN pada data Twitter dan mendapatkan hasil berupa akurasi dari sistem yang sudah ada sebesar 81% [3]. Penelitian selanjutnya dilakukan oleh Lopamudra Dey dan Sanjay Chakraborty melakukan analisis sentimen terkait review hotel dan film. Penelitian tersebut menggunakan klasifikasi Naïve Bayes dan KNN mendapatkan akurasi tertinggi dari Klasifikasi Naïve Bayes dengan dataset 4500 sebesar 82.43%.

Alasan peneliti memilih Twitter karena Twitter salah satu tempat untuk mendapatkan informasi dan menjadi salah satu media sosial yang sering digunakan masyarakat Indonesia. Selain peneliti ingin melakukan pengelompokan sentimen – sentimen masyarakat, peneliti juga ingin melakukan penerapan metode Naïve Bayes dalam melakukan analisis sentimen masyarakat terhadap UU Cipta kerja serta mengukur akurasi dari metode Naïve Bayes. Peneliti berharap dengan adanya penelitian ini dapat mengedukasi masyarakat dan menyediakan informasi terkait jumlah sentimen positif dan negatif terkait topik *Omnibus Law* UU Cipta Kerja dan memberikan bukti

empiris terkait penerapan metode Naïve Bayes dalam melakukan pendeteksian sentimen pada dataset Twitter.

II. METODE PENELITIAN

A. Analisis Sentimen

Analisis sentimen adalah proses yang dilakukan untuk menganalisis sebuah sentimen, opini masyarakat atau publik dalam bentuk teks yang digunakan untuk mengetahui emosi yang dirasakan masyarakat dalam sentimen kepada suatu objek seperti film, produk, acara, dan lain – lain [4]. Analisis sentimen memiliki nilai sentimen masing – masing sesuai dengan kecenderungan yang dikeluarkan masyarakat berupa opini atau emosi yang bernilai positif, negatif, atau netral [5]. Analisis sentimen merupakan teknik untuk memahami, mengklasifikasi, memahami dan mengidentifikasi pendapat terhadap sesuatu seperti objek dan lain – lain [6]. Analisis sentimen dapat dilakukan pada tingkat dokumen atau kalimat [7]. Analisis Sentimen memiliki beberapa tipe analisis yaitu:

1. *Fined-Grained Sentiment Analysis* : Sentimen analisis yang dilakukan dengan cara mengelompokkan beberapa opini atau sentimen masyarakat ke dalam beberapa nilai seperti positif, netral, dan negatif.
2. *Coarse-Grained Sentiment Analysis* : Sentimen analisis yang dilakukan dengan cara mengambil inti dari isi data sebagai sebuah sentimen atau opini yang memiliki sifat positif atau negatif [8].

Analisis Sentimen memiliki 5 langkah dalam prosesnya yaitu pengumpulan data, pemrosesan data, pendeteksian sentimen, klasifikasi sentimen, dan menganalisis sentimen atau presentasi output [9]. Tantangan

dalam melakukan analisis sentimen adalah beberapa kalimat yang memiliki dua makna berbeda antar positif atau negatif tergantung situasi [10].

B. Omnibus Law

Omnibus Law dapat diartikan yang dimulai dari kata Omnibus yang berasal dalam bahasa Latin yang berarti semuanya dan Law yang artinya hukum sehingga Omnibus Law adalah hukum untuk semua [11]. Omnibus Law adalah undang - undang yang mengatur berbagai macam materi berbeda dalam satu undang – undang [12]. Omnibus Law adalah undang – undang yang berfungsi untuk mengubah atau mencabut lebih dari satu undang – undang yang bertujuan untuk membenahi masalah – masalah karena peraturan yang tumpang tindih dan banyak sehingga berjalan dengan dinamis dengan masyarakat [13]. Omnibus Law digunakan oleh beberapa negara dengan sistem hukum *anglo saxon* atau *Common Law* seperti Irlandia, Inggris, Amerika Serikat, dan lain – lain untuk mengatasi regulasi dan aturan yang tumpang tindih sehingga tidak memakan waktu yang lama dan tidak mengeluarkan biaya secara berlebihan [14].

C. Twitter

Twitter adalah salah satu media sosial yang didirikan pada tahun 2006 yang berguna untuk melakukan proses mengirim dan membaca pesan yang sekarang dikenal sebagai cuitan atau *tweet*. Twitter merupakan salah satu media sosial yang sering digunakan masyarakat Indonesia untuk melakukan mengeluarkan pendapat atau opini kepada lembaga masyarakat, orang, dan lain – lain untuk memberikan kritik, saran, dan sebagainya [5]. Twitter adalah salah

satu layanan microblog yang memiliki 517 juta pengguna dengan pengguna Twitter di Indonesia yang mencapai 29,4 juta dan cuitan yang menggunakan Bahasa Indonesia mencapai 5,6% atau 19 juta cuitan per hari [15]. Pengguna dapat melakukan cuitan dengan batas panjangnya sebanyak 148 karakter di Twitter dan dapat membaca cuitan orang [16]. Salah satu bidang yang menggunakan media sosial twitter adalah bidang bisnis dalam melakukan pemasaran produk atau jasa, membuat relasi terhadap calon pelanggan atau pelanggan tetap, memberikan layanan kepada konsumen dan lain lain. Dataset yang digunakan dalam penelitian ini berupa data tweet yang dikumpulkan dari server twitter API.

D. Teks Preprocessing

Teks preprocessing adalah proses untuk membersihkan teks dari karakter yang mengganggu seperti link, emoji, username, dan karakter – karakter yang tidak digunakan untuk mendapatkan struktur teks yang baik [17]. Preprocessing data adalah kegiatan membersihkan data dari struktur yang buruk agar membuat data dapat memasuki tahap selanjutnya [18]. Preprocessing salah satu proses yang penting dalam melakukan *text mining* [19]. Proses *Preprocessing* yang tidak dilakukan dengan benar akan memberikan hasil yang menyimpang [20]. *Preprocessing* yang dipakai peneliti dalam penelitian ini adalah *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*.

1. Case Folding

Case Folding adalah proses untuk mengubah semua huruf dari a sampai z yang terdapat dalam dokumen menjadi huruf

kecil dan bila terdapat kata selain huruf akan dihilangkan sehingga dibaca delimiter. Dalam pemrosesan data, peneliti menggunakan *Case Folding* yang digunakan untuk mengurangi data yang berulang dan tidak konsisten sehingga mendapatkan hasil klasifikasi yang optimal [21].

2. Tokenisasi

Tokenisasi merupakan proses untuk melakukan pemisahan kata dalam sebuah kalimat atau teks. Tokenisasi melakukan pemecahan teks dari kalimat menjadi kata – kata yang berdiri sendiri – sendiri. Proses tokenisasi membantu dalam menghapus tanda baca seperti titik (.), koma (,), spasi, karakter angka yang ada pada kata tersebut, dan lain – lain [8]. Tokenisasi adalah proses untuk mendapatkan kata dengan melakukan pemisahan menggunakan bantuan spasi di antara kata [22].

3. Stopword Removal

Stopword merupakan daftar kata-kata yang tidak mempresentasikan isi dari suatu dokumen teks, *stopword* dilakukan untuk menghilangkan kata yang tidak memiliki arti. *Stopwords* berfungsi untuk menghapus kata penghubung karena dianggap tidak relevan seperti kata “dengan”, “untuk”, dan sebagainya yang tidak mengandung arti [18]. *Stopwords* menggunakan kamus sebagai acuan untuk menghapus kata yang tidak penting dan efisiensinya tergantung dari

daftar stopword yang telah dibuat [23].

4. Stemming

Stemming merupakan salah satu tahapan dari preprocessing yang berfungsi untuk mengubah kata yang memiliki imbuhan menjadi kata dasarnya dengan cara menghilangkan imbuhan awal atau imbuhan akhir kata tersebut [24]. Jenis – jenis imbuhan yaitu prefiks, infiks, sufiks, konfiks. Metode Stemming sudah dikembangkan di beberapa bahasa seperti Inggris, Prancis, Jerman, Belanda, Latin, Indonesia dan lain – lain [25]. Algoritma yang digunakan dalam dokumen berbahasa Indonesia adalah Algoritma Porter, Sastrawi, dan lain – lain [26].

E. Python

Python adalah salah satu bahasa pemrograman komputer yang *open source* sehingga mempermudah dalam pembelajaran, mendukung pemrograman berorientasi objek, mendukung *Internet of Things*, struktur kodingnya sesuai dengan prosedur, sederhana, dan lain – lain [27]. Python adalah bahasa yang memiliki sifat dinamis sehingga dapat menggunakan banyak metode [28]. Bahasa Pemrograman python dapat digunakan untuk membuat program desktop, Internet of Things, mobile, website, dan sebagainya. Python versi terbaru yang dipakai peneliti ketika menulis ini adalah 3.9.1. Python memiliki kelebihan dan fitur untuk membuat machine learning diantaranya :

- Memiliki struktur bahasa yang sederhana dan jelas
- Memiliki library yang luas dan banyak.
- *Open source*
- Memiliki panjang kode yang lebih sedikit dibanding pemrograman lainnya sehingga mengirit waktu dalam proses pengerjaan.

F. Naïve Bayes

Naive Bayes adalah sebuah algoritma klasifikasi untuk menghitung kumpulan dari probabilitas yang terjadi dengan menghitung frekuensi dan kombinasi nilai dari kumpulan data yang akan digunakan [29]. Naïve Bayes adalah salah satu metode klasifikasi dalam memprediksi objek yang belum diketahui berdasarkan data sebelumnya menggunakan statistik dan peluang [30]. Keuntungan menggunakan klasifikasi Naïve Bayes adalah membutuhkan dataset yang sedikit untuk proses klasifikasi [31]. Klasifikasi Naive Bayes menganggap fitur tertentu dalam sebuah kelas tidak terikat atau terkait dengan yang lainnya. Rumus teorema Naïve Bayes seperti pada gambar 2.1 berikut dengan keterangannya [32].

$$P(C | X) = \frac{P(X | C) P(C)}{P(X)}$$

Gambar 2.1 Rumus Naïve Bayes

Keterangan :

- *c* : Hipotesis data
- *X* : Data class yang belum diketahui
- *P(c |X)* : Probabilitas hipotesis *c* berdasarkan kondisi dari data *X*

- $P(X|c)$: Probabilitas kondisi data X pada hipotesis C
- $P(X)$: Probabilitas data

F. Kerangka Berpikir



Gambar 2.2 Kerangka Berpikir

Penelitian ini dilakukan peneliti dimulai dengan melihat banyak sentimen masyarakat yang pro dan kontra terkait UU Cipta Kerja di berbagai media sosial salah satunya Twitter. Tahap awal yang dilakukan dalam penelitian ini adalah mengumpulkan dataset dengan melakukan *crawling* data menggunakan Twitter API. Peneliti membutuhkan *credential keys* untuk dapat melakukan proses *crawling* dengan meminta akses kepada pengembang twitter. Dataset yang telah dikumpulkan dari *crawling* di Twitter API akan dilakukan pelabelan data sesuai dengan kategori sentimen tersebut yaitu sentimen positif atau sentimen negatif. Data yang sudah dilabeli akan dilakukan preprocessing untuk membersihkan slang atau kata yang tidak yang tidak diperlukan untuk mengoptimalkan proses klasifikasi dan meningkatkan akurasi. Proses preprocessing dapat dilihat pada gambar 2.3.



Gambar 2.3 Proses *Preprocessing*

Preprocessing data dalam penelitian ini menggunakan 5 proses. Proses pertama adalah *Cleansing* data untuk menghilangkan username, link, *hashtag*, angka, tanda baca, dan menghilangkan tulisan retweet atau RT. Proses kedua adalah proses *Case*

Folding untuk mengecilkan semua huruf, proses ketiga adalah proses *stopword removal* untuk menghilangkan kata hubung seperti di, yang, ke, dan kata penghubung lainnya. Proses keempat adalah proses *Stemming* untuk mengubah kata yang memiliki imbuhan menjadi kata dasarnya dan proses kelima yaitu tokenisasi untuk memecahkan sebuah kalimat menjadi kata. Data yang sudah dilakukan *preprocessing* dibagi menjadi 2 data yaitu data latih dan data uji.

Kedua data tersebut akan dilakukan *CountVectorizer* sehingga memudahkan dalam proses klasifikasi menggunakan Naïve Bayes. Tahap selanjutnya adalah melakukan pembelajaran terhadap mesin dengan menggunakan data latih dengan metode *Naïve Bayes*. Setelah dilakukan tahap pembelajaran akan dilakukan pengujian dengan menggunakan data uji. Hasil yang didapatkan dalam tahapan tersebut berupa akurasi, *precision*, *f-measure*, dan *recall* mesin dalam melakukan prediksi pada data uji dan mendapatkan hasil klasifikasi yang akan menjadi model klasifikasi untuk melakukan prediksi dari dataset baru di luar data yang telah dikumpulkan sebelumnya yang hasilnya berupa prediksi sentimen dari dataset yang belum dilabeli.

G. Metode Penelitian

Dalam tahapan ini peneliti membutuhkan dataset berupa cuitan yang sudah dikumpulkan untuk membuat logika model pembelajaran mesin terkait klasifikasi sentimen UU Cipta Kerja. Data yang sudah dikumpulkan akan dilakukan *preprocessing* untuk menghindari data menjadi tidak akurat atau relevan sehingga hasil prediksi dapat

dipertanggungjawabkan. Setelah dilakukan *preprocessing*, data akan diubah menjadi vektor untuk memudahkan dalam menganalisis teks dan membuat model pembelajaran mesin.

Peneliti menggunakan klasifikasi Multi- nomial Naïve Bayes dalam pengembangan model dalam penelitian ini untuk melakukan klasifikasi sentimen. Data dikumpulkan menggunakan metode *crawling* data dengan Twitter API yang disimpan dalam format CSV, Twitter API adalah bagian dari Twitter developer dengan bentuk program untuk memudahkan pengembang lain dapat mengakses informasi yang terdapat dalam Twitter. Dalam melakukan *crawling* data dibutuhkan *credential keys* berupa *token* dan *key* untuk mengakses Twitter API. Dalam mendapatkan *credential keys* dibutuhkan permintaan akses data kepada pengembang twitter untuk proses verifikasi. Verifikasi yang dilakukan memiliki sedikit hambatan karena waktu yang dibutuhkan sedikit lama untuk mendaftarkan dan melakukan verifikasi akun dengan menanggapi pertanyaan – pertanyaan yang diberikan pengembang lewat email yang sudah didaftarkan dan harus memiliki respon yang tanggap dalam menanggapi pertanyaan – pertanyaan tersebut.

Data diambil peneliti pada April 2020 dan banyak data yang didapatkan 300 cuitan. Data yang terkumpul dilakukan *labelling* manual berdasarkan sentimen yang mendukung dan tidak terkait pengesahan UU Cipta Kerja. Peneliti mendapatkan total 160 cuitan setelah dilakukan proses *labelling*. Dataset yang dikumpulkan masih berupa data

noise yang memiliki simbol, link, kata hubung dan lain –

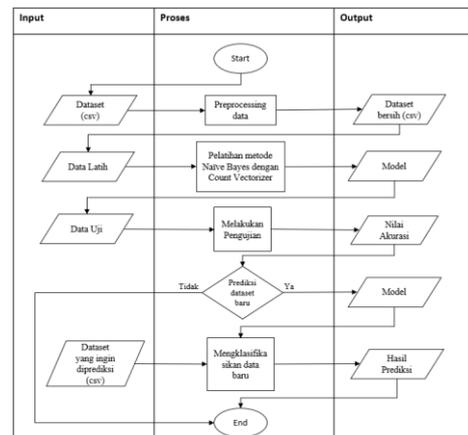
lain sehingga dilakukan *preprocessing*. *Preprocessing* yang dilakukan diantaranya adalah penghapusan karakter spasi, link, gambar, RT, membuat semua huruf menjadi kecil, membuang kata yang tidak memiliki bobot atau penting, mengubah kata yang memiliki imbuhan menjadi kata dasar dan dilakukan pemisahan kata untuk memudahkan dalam proses klasifikasi. Data yang sudah dibersihkan disimpan kembali dalam format CSV.

Data yang sudah disimpan, peneliti menggunakan library *Scikit-learn CountVectorizer* untuk mengubah angka menjadi vektor sehingga memudahkan proses klasifikasi menggunakan metode Naïve Bayes dan hasilnya membuat model pembelajaran mesin yang membantu dalam proses analisis sentimen UU Cipta Kerja. Dalam penelitian ini didapatkan sebanyak 19 cuitan dari total 22 cuitan untuk sentimen positif yang di prediksi benar oleh mesin dan sebanyak 22 cuitan dari total 26 cuitan sentimen negatif yang dijawab dengan benar. Perangkat keras dan perangkat lunak yang digunakan peneliti dalam penelitian ini dapat dilihat pada tabel 3.1.

Tabel 3.1 Instrumen Penelitian

Pada tahapan ini, dilakukan pembuatan desain alur dari logika model. Pertama, peneliti melakukan *preprocessing* dari dataset yang dikumpulkan dari *crawling* di twitter terkait UU Cipta Kerja dan sudah diberikan label sentimen secara manual untuk membantu dalam proses melatih data. Proses *preprocessing* meliputi pembersihan kata dari simbol, link, angka, username, *hashtag*, *double* spasi, menghilangkan kata hubung, menghilangkan imbuhan dan merubah kata menjadi kata dasarnya dan melakukan pemecahan kalimat jadi beberapa kata. Data yang telah melewati proses *preprocessing* akan disimpan ke dalam format CSV. Setelah itu, peneliti membagi data ke dalam 2 bagian yaitu data latih dan data uji.

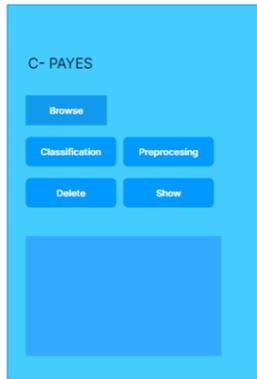
Data tersebut dalam penelitian ini akan diubah menjadi vektor untuk memudahkan proses klasifikasi. Pelatihan mesin dalam penelitian ini menggunakan klasifikasi Naïve Bayes. Hasil pengujian yang didapatkan berupa *precision*, *recall* dan akurasi. Hasil analisis sentimen dalam penelitian ini menggunakan *classification report* dan *confusion matrix* untuk dilakukan evaluasi. Model yang sudah didapatkan dari proses pembelajaran akan digunakan untuk melakukan prediksi dataset baru di luar dari dataset yang di *crawling* sebelumnya. Hasil dari prediksi tersebut berupa pelabelan sentimen secara otomatis oleh mesin. Alur logika model dapat dilihat pada gambar 2.4.



Gambar 2.4 Desain Alur Logika Model

Pada tahapan ini, peneliti melakukan rancangan desain antarmuka untuk memudahkan pengguna dalam menggunakan aplikasi dan mendapatkan informasi terkait analisis sentimen UU Cipta kerja secara keseluruhan.

Perangkat keras	Perangkat lunak
Laptop dengan sistem operasi Windows 10 Home Single Language 64-bit, Processor i7-865U, RAM 8GB, Nvidia MX 230, 500GB Storage	Python 3.8.5, Sypder, Qt Designer, Numpy, Sklearn, Sastrawi, Pandas, Nltk, Re, PyQt5, Tweepy



Gambar 2.5 Desain antarmuka

III. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini berupa model klasifikasi untuk membantu dalam proses analisis sentimen UU Cipta Kerja. Pada tahap awal mengumpulkan sentimen dari Twitter dan melakukan pelabelan secara manual. Penelitian menggunakan API twitter untuk membantu dalam proses *crawling*. Dataset yang berhasil dikumpulkan dan diberi label adalah 160 cuitan dengan 86 cuitan bersentimen negatif dan 74 cuitan bersentimen positif. Setelah itu dilakukan *preprocessing* data dengan menggunakan *Case Folding*, *Stopword Removal*, *Stemming*, dan tokenisasi dan hasil data setelah disimpan ke dalam CSV dapat dilihat pada gambar 3.1

```
[perkara,'kriminalisasi','empat','mahasiswa','tolak','omnibus','law','bem','undip','bem','th','undip']
[pokok,'ngab','bijak','kata','titah','jokowi','maha','benar','bipang','ka','corona']
[sajak,'sama','kaya','jmn','omnibus','law','jauh','hari','sosialisasi','k','kampus','yg','ngikuti','dikit','skt']
[bima,'anya','keluh','omnibus','law','bikin','pemda','hilang','yatribusi']
[wow,'keren','habis','kungfu','satu','satu','jangan','smpe','kita','sesal','kemudian','jika','sudah']
[agus,'haris','desak','pemkot','bontang','inventarisasi','perda','dampak','uu','omnibus','law','cipta','kerja']
[demo,'tolak','laku','bem','ui','lantar','uu','ciptaker','nilai','regulasi','provokatif']
[curhat,'menteri','investasi','wali','kota','indonesia','ungkap','khawatir','soal','omnibus','law']
[bem,'ui','demo','tolak','omnibus','law','bukan','postingan','jumbuh','hidayat']
[press,'release','halo','sobat','km','omnibus','law','gandeng','status','genting','terus','kawal','lanjut']
```

Gambar 3.1 Teks Preprocessing

Kemudian data tersebut dilakukan proses pelabelan secara manual menurut kecenderungan sentimen masyarakat. Pelabelan dengan menulis angka 1 untuk sentimen positif dan angka 2 untuk sentimen negatif.

```
[perkara,'kriminalisasi','empat','mahasiswa','tolak','omnibus','law','bem','undip','bem','th','undip']
[pokok,'ngab','bijak','kata','titah','jokowi','maha','benar','bipang','ka','corona']
[sajak,'sama','kaya','jmn','omnibus','law','jauh','hari','sosialisasi','k','kampus','yg','ngikuti','dikit','skt']
[bima,'anya','keluh','omnibus','law','bikin','pemda','hilang','yatribusi']
[wow,'keren','habis','kungfu','satu','satu','jangan','smpe','kita','sesal','kemudian','jika','sudah']
[agus,'haris','desak','pemkot','bontang','inventarisasi','perda','dampak','uu','omnibus','law','cipta','kerja']
[demo,'tolak','laku','bem','ui','lantar','uu','ciptaker','nilai','regulasi','provokatif']
[curhat,'menteri','investasi','wali','kota','indonesia','ungkap','khawatir','soal','omnibus','law']
[bem,'ui','demo','tolak','omnibus','law','bukan','postingan','jumbuh','hidayat']
[press,'release','halo','sobat','km','omnibus','law','gandeng','status','genting','terus','kawal','lanjut']
[hidat,'beni','pagi','sudah','meeting','nggak','omnibus','law','regulasi','kay']
[halo,'gara','omnibus','law','jmn','bembur','jrt','maliki','panjang']
[pejabat,'polisi','republik','indonesia','beni','haki','munandar','kemplang','uang','akyat','halu','uu','tahu']
[pejabat,'ibahi','haki','hajar','hidup','layak','meerasi','pejabat','cipta','beror','serang','alim']
[malai,'anak','tmi','mahasiswa','buruh','semak-email','turun','jaka','susu','huntaf','batal','uu','omnibus']
[sidi,'ambivert','mau','ngadukan','polisi','lha','kari','memang','benar','porsi','weewenang','uc','omnibus']
[nyata,'manfaat','omnibus','law','tanggap','kasa','orang','china','sukses','banda','cipta','kerja','nyata','china']
[dpr,'pejabat','baru','bahas','uu','kata','ulang','tembaga','uang','agustus','september']
[berak,'herus','puark','ekonomi','akyat','akibat','semenah','pihak','jrt','kni','lmbas','omnibus','law']
[berita,'soal','luhuf','banga','omnibus','law','biayi','pegawai','bumi','kasa','virus','corona','tanggung','hingga','rp','jda']
[saat,'izin','operasi','satu','satu','pejabat','nomor','tahun','selenggara','hidang']
[aktiva,'jumbuh','hidayat','jalan','sidang','kasus','duga','sebar','berita','bohong','adil','seger']
```

Gambar 3.2 Pelabelan Secara Manual

Peneliti melanjutkan dengan melakukan pemisahan data latih dan uji dengan presentasi 70% dan 30%. Kemudian data tersebut, dilakukan proses pelatihan mesin menggunakan metode Naïve Bayes dengan bantuan CountVectorizer. Pengujian dilakukan setelah mesin berlatih dengan data uji sebesar 30% atau 48 cuitan. Setelah itu hasil pengujian ditampilkan dalam bentuk *confusion matrix* dan *classification report*.

```
[[19 3]
 [ 4 22]]
```

Gambar 3.3 Confusion Matrix

Hasil dari *confusion matrix* dengan angka 19 adalah *True Positive* (TP) yang artinya data dengan sentimen positif yang di prediksi dengan benar oleh mesin sebanyak 19 sentimen, angka 4 dalam *confusion matrix* adalah *False Positive* (FP) yang artinya data dengan sentimen positif yang dijawab salah oleh mesin sebanyak 4 sentimen, angka 3 dalam *confusion matrix* adalah *False Negative* (FN) yang artinya data dengan sentimen negatif yang dijawab salah oleh sistem sebanyak 3 sentimen, dan angka 22 dalam *confusion matrix* adalah *True Negative* (TN) yang artinya data dengan sentimen negatif yang dijawab benar oleh sistem sebanyak 22 sentimen. *Confusion matrix* yang telah didapatkan dapat menghitung akurasi, *recall*, dan *precision* secara manual.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} = \frac{19}{19 + 4}$$

Gambar 3.4 Perhitungan Akurasi Manual

Pada gambar 3.4 adalah cara mendapatkan akurasi secara manual menggunakan TF-IDF dari *confusion matrix* yang didapatkan dengan menggunakan *True Positive* ditambah dengan *True Negative* dibagi dengan total cuitan yang digunakan.

$$\text{Akurasi} = \frac{19 + 22}{19 + 4 + 3 + 22} = 0.854$$

Pada perhitungan yang dilakukan penelitian ini secara manual untuk menghitung akurasi sebesar 0.854 atau 85%.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{atau} \quad \frac{\text{TN}}{\text{TN} + \text{FN}}$$

Gambar 3.5 Perhitungan Precision Manual

Gambar 3.5 adalah cara mendapatkan *precision positive* secara manual menggunakan TF-IDF dari *confusion matrix* yang didapatkan dengan menggunakan *True Positive* (TP) dibagi dengan pertambahan *True Positive* (TP) dengan *False Positive* (FP). *Precision negative* secara manual menggunakan TF-IDF dari *confusion matrix* yang didapatkan dengan menggunakan *True Negative* (TN) dibagi dengan pertambahan *True Negative* (TN) dengan *False Negative* (FL).

$$\begin{aligned} \text{Precision Positive} &= \\ &= 0.826 / 0.83 \end{aligned}$$

$$\begin{aligned} \text{Precision Negative} &= \\ &= 0.88 \end{aligned}$$

$$\begin{aligned} \text{Rata - rata Precision} &= (0.83 + \\ &0.88) / 2 = \\ &0.855 \end{aligned}$$

Penelitian ini *precision* yang didapatkan *precision* positif dan *precision* negatif secara berturut – turut adalah 0.83 dan 0.88. Rata – rata *precision* dari perhitungan manual adalah 0.855. Akurasi, *recall*, dan *precision* yang didapatkan dalam penelitian ini dapat dilihat secara otomatis dengan menggunakan *classification report* seperti pada gambar 3.6.

	precision	recall	f1-score	support
1	0.83	0.86	0.84	22
2	0.88	0.85	0.86	26
accuracy			0.85	48
macro avg	0.85	0.85	0.85	48
weighted avg	0.86	0.85	0.85	48

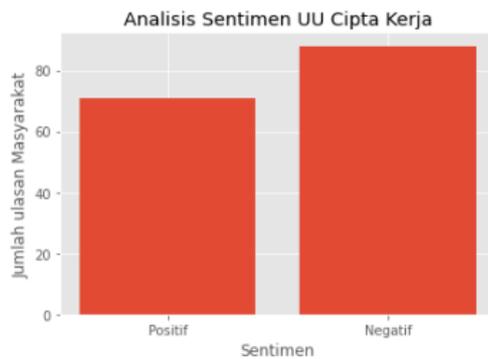
Gambar 3.6 Classification Report

Hasil dari klasifikasi dalam penelitian ini mendapatkan akurasi sebesar 0.85 atau 85%, *recall* 0.85, dan *precision* sebesar 0.85 yang didapatkan secara manual atau otomatis. Sentimen positif yang di prediksi dengan benar dalam penelitian ini sebanyak 19 cuitan dari total 22 cuitan untuk sentimen positif dan sebanyak 22 cuitan dari total 26 cuitan yang di prediksi dengan benar oleh mesin untuk sentimen negatif.

['pak', 'jokowi', 'anda', 'kepala', 'negara', 'laku', 'hubung', 'toksik', 'rakyat', 'ikut', 'bunuh', 'rakyat']	Negatif
['omnibus', 'law', 'cipta', 'kerja', 'taju', 'buka', 'lapang', 'kerja', 'baru', 'mudah', 'buka', 'usaha', 'berant']	Positif
['susanti', 'dukung', 'omnibus', 'law']	Positif
['sahkan', 'omnibus', 'law']	Positif
['demo', 'terjadi', 'tolak', 'omnibus', 'law']	Negatif
['mau', 'melawan', 'siapa']	Negatif
['omnibus', 'law', 'buat', 'mepet', 'waktu', 'pemilu']	Negatif
['omnibus', 'law', 'cipta', 'kerja', 'sekarang', 'investor', 'lirik', 'indonesia']	Positif
['tolak', 'omnibus', 'law', 'pemuda']	Negatif
['indonesia', 'promosi', 'omnibus', 'law', 'luar', 'negeri']	Positif
['demo', 'lawan', 'omnibus', 'law']	Negatif
['omnibus', 'law', 'tumpang', 'tindih', 'aturan']	Negatif
['omnibus', 'law', 'investasi', 'cipta', 'lapang', 'kerja']	Positif
['produktif', 'indonesia', 'tingkat', 'omnibus', 'law']	Positif
['aneh', 'teriak', 'pemerintah', 'fasistlah']	Negatif
['omnibus', 'law', 'mudah', 'lindung', 'daya', 'umkm', 'koperasi', 'investasi']	Positif
['teriak', 'terus', 'puru', 'ekonomi', 'rakyat', 'akibat', 'semenah', 'pihak', 'pt', 'ini', 'imbas', 'omnibus']	Negatif
['aktifis', 'jumbuh', 'hidayat', 'jalan', 'sidang', 'kasus', 'duga', 'sebar', 'berita', 'bohong', 'adil', 'negeri']	Positif

Gambar 3.7 Hasil Analisis Sentimen

Gambar 3.7 menunjukkan hasil analisis sentimen yang didapatkan oleh peneliti dalam penelitian ini. Diagram yang didapatkan dalam penelitian ini seperti yang dapat dilihat pada gambar 3.8 dengan perolehan sentimen positif sebanyak 74 cuitan dan untuk sentimen negatif sebanyak 86.



Gambar 3.8 Diagram Analisis sentimen

Dalam data pada gambar 3.9 adalah *True Positive* karena hasil prediksi terklasifikasi positif dan di prediksi benar oleh sistem. Kata – kata yang menjadi bobot dalam membantu sistem untuk membaca kalimat positif adalah ekonomi, positif, sejahtera, dan lain – lain.

Tweet Jawaban	Prediksi
omnibus law dongkrak ekonomi indonesia	Positif
omnibus law uu cipta kerja beri manfaat positif nelayan	Positif
pasti izin operasi acu atur pemerintah nomor tahun selenggara bidang	Positif
jangnan rusuh terkait omnibus law	Positif
omnibus law cipta kerja beri dampak positif cipta lapangan kerja baru	Positif
uu cipta kerja bawa efek positif pasar modal	Positif
jerman positif reformasi indonesia omnibus law	Positif
omnibus law positif indonesia	Positif
produktif indonesia tingkat omnibus law	Positif
omnibus law dampak positif bagi indonesia timur	Positif
bawa investor asing aseng indonesia tolak alasan rasist	Positif
investor singapura positif omnibus law	Positif
efek omnibus law ekonomi bergerak positif	Positif
obi maju pariwisata ni	Positif
umkm positif hadir uu cipta kerja	Positif
stop provokasi omnibus law cipta kerja dukung bawa dampak positif	Positif
luck sejahtera petani	Positif
indonesia promosi omnibus law luar negeri	Positif
snitex positif uu omnibus law cipta kerja	Positif

Gambar 3.9 True Positive

Dalam data pada gambar 3.10 adalah *True Negative* karena hasil prediksi terklasifikasi negatif dan di prediksi benar oleh sistem. Kata – kata yang menjadi bobot dalam membantu sistem untuk membaca kalimat negatif adalah tka, tolak, resah, china, dan lain – lain.

Tweet Jawaban	Prediksi
nyata manfaat omnibus law sangat rasa orang china sukses anda cipta kerja nyata china	Negatif
polkongnya bijak kata tka tka Jokowi maha benar bipang tka corona	Negatif
kayak kasus anak smk kmeren yg pegang bendera pas demo omnibus law	Negatif
mau melawan siapa	Negatif
omnibus law diem masyarakat	Negatif
tolak omnibus law pemuda	Negatif
omnibus law lemah kpk	Negatif
china masuk n buruh taju omnibus law	Negatif
sudah ikut bunuh rakyat Kalimantan lalu bangri ikut sahkan omnibus law pak Jokowi	Negatif
sajak sama kaya jrn omnibus law jauh hasi sosialisasi k kampus yg ngkritik dikit sdi	Negatif
wahai Jokowi sang lelaki dusta tebus dosa buki diri sudah ikut kesah omnibus law	Negatif
rikel kasih china makanya bondong kesni sama emas tembaga di kasih	Negatif
klo bijak Jokowi revisi uu kpt uu o	Negatif
lawan omnibus law tolak	Negatif
dampak revisi uu kpt protes banyak orang tak gabris tinggal tunggu dampak omnibus law	Negatif
omnibus law tka asing masuk	Negatif
waktu buruh mahasiswa demo dntah soal batal uu omnibus law mana kalkan semua bangsa kalkan sekolahan	Negatif
pak Jokowi ikut buruh rakyat yg tewas banji sebut kesah omnibus law yg mudah	Negatif
rapat paripurna bolos omnibus law digagal in koraptor rajaJela suaw yo males mbayar pajak	Negatif
nyala omnibus law makmur gelmir rakyat indonesia	Negatif
kpa kalo omnibus law jalan sos yg anti weseng tzn dnterit udah ga yg biasa main proyek	Negatif

Gambar 3.10 True Negative

Dalam data pada gambar 3.11 adalah *False Negative* karena hasil prediksi terklasifikasi negatif namun di prediksi salah oleh sistem. Kata – kata yang menjadi bobot dalam membantu sistem untuk membaca kalimat negatif adalah corona, layar, dan lain – lain selain itu karena dataset yang digunakan peneliti sedikit sehingga terjadi kesalahan prediksi.

Tweet Jawaban	Prediksi
berita soal luhut bangga omnibus law biaya pegawai bumi kena virus corona tanggung hingga Rp juta	Positif
uu layar omnibus law	Negatif
nah loh lali klo buruh yg demo bayar cuman rusuhin negara aja duga demo omnibus	Negatif

Gambar 3.11 False Negative

Dalam data pada gambar 3.12 adalah *False Negative* karena hasil prediksi terklasifikasi positif namun di

prediksi salah oleh sistem. Kata – kata yang menjadi bobot dalam membantu sistem untuk membaca kalimat negasi positif adalah omnibus law, mk, regulasi dan lain – lain selain itu karena dataset yang digunakan peneliti sedikit sehingga terjadi kesalahan prediksi.

	Tweet Jawaban	Prediksi
	omnibus law buat mepet waktu pemilu Negatif	Positif
	omnibus law bawa mk Negatif	Positif
	tidur dari pagi udah meeting ngulik omnibus law regulasi tka Negatif	Positif

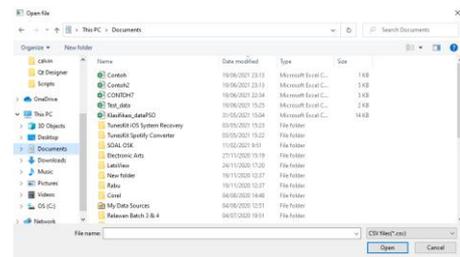
Gambar 3.12 False Positive

Hasil dari desain penelitian ini adalah aplikasi analisis sentimen yang sudah dirancang dengan menggunakan QtDesigner untuk antarmuka. Peneliti membangun berdasarkan desain yang sudah dibuat dan disesuaikan dengan kebutuhan dari penelitian yang mudah untuk digunakan. Seperti yang ditunjukkan pada Gambar 3.13.



Gambar 3.13 Hasil Desain Implementasi

Peneliti membuat tampilan yang terdiri dari 5 tombol dan 1 tabel view. Tombol – tombol terdiri dari *Browse*, *Pre-Processing*, *Classification*, *Delete*, dan *Show*. Tombol *Browse* berfungsi untuk membuka jendela *file* untuk memilih *file* terkait UU Cipta Kerja dan akan dilakukan proses prediksi dan klasifikasi.



Gambar 3.14 Jendela Open File

Tombol *Pre-processing* melakukan *preprocessing* data yang masih memiliki simbol, angka, link, username, dan lain – lain yang masih *noise* untuk membantu dalam proses klasifikasi UU Cipta Kerja sehingga data yang didapatkan akan relevan dan dapat dipertanggungjawabkan. Contoh teks yang sudah dilakukan proses *preprocessing* dapat dilihat pada gambar 3.15

['ya', 'omnibus', 'law', 'mudah', 'rusak', 'alam', 'akibat', 'jadi', 'sering', 'banjir', 'di
 ['ya', 'jokowi', 'khianat', 'negara', 'indonesia', 'karwna', 'kesah', 'omnibus', 'law']
 ['pak', 'jokowi', 'bukan', 'lindung', 'alam', 'nyawa', 'rakyat', 'maupun', 'buruh', 'si
 ['wahai', 'jokowi', 'sang', 'lelaki', 'dusta', 'tebus', 'dosa', 'bukti', 'diri', 'sudah', 'iku
 ['sudah', 'ikut', 'bunuh', 'rakyat', 'kalimantan', 'lalu', 'banjir', 'ikut', 'sahkan', 'om
 ['oke', 'jauh', 'bukti', 'jelas', 'pak', 'jokowi', 'ikut', 'bunuh', 'rakyat', 'kalimantan', '
 ['pak', 'jokowi', 'pernah', 'dusta', 'banjir', 'kalimantan', 'curah', 'hujan', 'padahal',
 ['ini', 'bukti', 'omnibus', 'law', 'sebab', 'banjir', 'pembunub', 'rakyat', 'kalimantan

Gambar 3.15 Teks Preprocessing

Tombol *Delete* berfungsi untuk menghilangkan data atau mengosongkan dataset yang sudah dimasukkan ke dalam aplikasi tersebut seperti pada gambar 3.16. Tombol *Delete* digunakan ketika salah memasukkan dataset, terjadi klasifikasi error, dan lain sebagainya sehingga tidak diperlukan untuk mengulang menjalankan program kembali.



Gambar 3.16 isi data yang dihapus

Tombol *Show* berfungsi untuk menampilkan dataset yang sudah dimasukkan ke dalam aplikasi tersebut seperti pada gambar 3.17. Tombol *Show* digunakan ketika peneliti ingin menampilkan data sebelum dilakukan proses *preprocessing* atau klasifikasi dan data sesudah dilakukan proses *preprocessing* dan klasifikasi.



Gambar .3.17 Menampilkan Data

Tombol *Classification* berfungsi untuk melakukan analisis sentimen dari dataset yang sudah dimasukkan ke dalam aplikasi tersebut seperti pada gambar 3.18. Pada gambar 3.18, dataset berhasil di klasifikasi menggunakan Klasifikasi Naïve Bayes dan memprediksi dataset berupa teks dengan sentimen yang

sudah di prediksi berupa sentimen positif atau sentimen negatif.



Gambar 4.14 Menampilkan Data Klasifikasi

IV. SIMPULAN

Simpulan dari penelitian ini dalam melakukan analisis sentimen terhadap UU Cipta kerja serta pembuatan aplikasi sebagai berikut:

1. Jumlah sentimen negatif yang didapatkan peneliti sebanyak 86 cuitan dan sentimen positif sebanyak 74 cuitan.
2. Penelitian ini menggunakan 160 cuitan dengan menggunakan data uji sebesar 48 data dengan sentimen positif yang di prediksi dengan benar oleh mesin sebanyak 19 cuitan dari total 22 cuitan untuk sentimen positif yang dan sebanyak 22 cuitan dari total 26 cuitan yang di prediksi oleh mesin dengan sentimen negatif.
3. Meminta akses data dalam melakukan *crawling* data Twitter API menjadi sedikit hambatan karena menunggu proses verifikasi dan proses verifikasi yang terlalu banyak

sehingga membutuhkan waktu untuk dilakukan dan harus memiliki responsif yang tanggap menanggapi pertanyaan yang akan diberikan *developer* lewat email.

4. Akurasi yang didapatkan dalam penelitian ini sebesar 85% dengan rata – rata precision, recall, dan f1-score sebesar 85%. Dengan rata – rata akurasi 85% bisa dikatakan lumayan bagus dan data yang di prediksi dapat dipertanggungjawabkan.
5. Menggunakan dataset yang lebih banyak untuk meningkatkan besar akurasi
6. Mengkombinasikan Naïve Bayes dengan Particle Swarm Optimization sebagai *feature selection* untuk meningkatkan akurasi.
7. Menggunakan Support Vector Machine dengan Particle Swarm Optimization untuk meningkatkan keakuratan prediksi.

DAFTAR RUJUKAN

- [1] A. Suntoro, "Implementasi Pencapaian Secara Progresif dalam Omnibus Law Cipta Kerja," *Jurnal Ham*, vol. 12, pp. 1-18, 2021.
- [2] "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," *Procedia Computer Science*, vol. 127, pp. 511-520, 2018.
- [3] Soudamini Hota, Sudhir Pathak, "KNN classifier based approach for multi-class sentiment," *International Journal of Engineering & Technology*, vol. 7, pp. 1372-1375, 2018.
- [4] Pooja Mehta, Dr.Sharnil Pandya, "A Review On Sentiment Analysis Methodologies,," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, pp. 601 - 609, 2020.
- [5] F.Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter," *INVOTEK POLBENG*, vol. 3, p. 50, 2018.
- [6] Ms. Gaurangi Patil, Ms. Varsha Galande, Mr. Vedant Kekan, Ms. Kalpana Dange, "Sentiment Analysis Using Support Vector," *International Journal of Innovative Research in Computer*, vol. 2, pp. 2607-2612, 2014.
- [7] E. M. Younis, "Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study," *International Journal of Computer Applications*, vol. 112, pp. 44-48, 2015.
- [8] Lian Ardiani, Herry Sujaini, Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, p. 183, 2020.
- [9] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo, "Approaches, Tools and Applications for Sentiment," *International Journal of Computer Applications*, vol. 125, pp. 26-33, 2015.
- [10] Vidisha M. Pradhan, Jay Vala, Prem Balani, "A Survey on Sentiment Analysis Algorithms for Opinion Mining," *International Journal of Computer Applications*, vol. 133, pp. 7-11, 2016.
- [11] A. Fitryantica, "Harmonisasi Peraturan Perundang-Undangan Indonesia melalui," *Jurnal Gema Keadilan*, vol. 6, pp. 300-316, 2019.
- [12] I Putu Eka Cakra, Aditya Yuli Sulistyawan, "Kompabilitas Penerapan Konsep Omnibus Law Dalam Sistem Hukum Indonesia," *Jurnal CREPIDO*, vol. 02, pp. 59-69, 2020.
- [13] A. Putra, "Penerapan Omnibus Law dalam Upaya Reformasi Regulasi," *Jurnal Legislasi Indonesia*, vol. 17, pp. 1-10, 2020.
- [14] S. S. Kukuh Tejomurti, "The Critical Study of the Omnibus Bill on Job Creation Based on John Rawls View on Justice," *UNNES LAW JOURNAL*, vol. 6, pp. 188-204, 2020.

- [15] Lutfie Harrie, Hidayat Rahmat, Cahyani Leni, "AN INVESTIGATION OF IMPLEMENTATION SAMSUNG PROMOTION ON MEDIA (YOUTUBE, TWITTER AND FACEBOOK) FOR PURCHASING DECISION," *6th International Conference on Evolving Trends in Social Sciences and Business Studies*, 2017.
- [16] Varsha Sahayak, Vijaya Shete, Apashabi Pathan, "Sentiment Analysis on Twitter Data," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 2, no. 1, pp. 178-183, 2015.
- [17] Rizki Aulianita, Achmad Rifai, "Optimasi Particle Swarm Optimization Pada Naive Bayes Untuk Sentiment Analysis Furniture," *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS*, vol. 3, pp. 31-40, 2018.
- [18] Yuris Alkhalifi, Windu Gata, Arfhan Prasetyo, Imam Budiawan, "Analisis Sentimen Penghapusan Ujian Nasional pada Twitter Menggunakan Support Vector Machine dan Naive Bayes berbasis Particle Swarm Optimization," *Jurnal CoreIT*, vol. 6, pp. 71-78, 2020.
- [19] Dr. S. Vijayarani, Ms.J. Ilamathi, Ms.Nithya, "Preprocessing Techniques for Text Mining - An Overview," *International Journal of Computer Science & Communication Networks*, vol. 5, pp. 7-16, 2015.
- [20] Mohammed Zuhair Al-Taie, Seifedine Kadry, Joel Pinho Lucas, "Online data preprocessing: a case study approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, pp. 2620 - 2626, 2019.
- [21] Hidayatul Ma'rifah, Aji Prasetya Wibawa, Muhammad Iqbal Akbar, "Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing," *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, vol. 2, pp. 70-78, 2020.
- [22] Hamimah Alatas Hendri Murfi, Alhadi Bustamam, "Topic Detection using Fuzzy C-Means with," *Internasional Journal Advance Soft Compu. Appl*, vol. 10, pp. 206-222, 2018.
- [23] Jaideepsinh K. Raulji, Jatinderkumar R. Saini, PhD, "Stop-Word Removal Algorithm and its Implementation," *International Journal of Computer Applications*, vol. 150, pp. 15-17, 2016.
- [24] Hadiyatun Najjichah, Abdulk Syukur, Hendro Subagyo, "PENGARUH TEXT PREPROCESSING DAN KOMBINASINYA PADA PERINGKAS DOKUMEN OTOMATIS TEKS BERBAHASA INDONESIA," *Jurnal Teknologi Informasi*, vol. 15, pp. 1-11, 2019.
- [25] Abdulfattah Omar, Wafya Ibrahim Hamouda, "The Effectiveness of Stemming in the Stylometric," *International Journal of Advanced Computer Science and Applications*, vol. 11, pp. 116-121, 2020.
- [26] L. F. Narulita, "Pengaruh Proses Stemming Pada Kinerja Analisa Sentimen Pada Review Buku," *Hasil Penelitian LPPM Untag Surabaya*, vol. 03, pp. 55-59, 2018.
- [27] M. Lutz, Python Pocket Reference, United States of America: O'REILLY Media, 2014.
- [28] J. O. Knowlton, Python Create - Modify - Reuse, Canada: Wiley Publishing, 2008.
- [29] Mucahid Mustafa Saritas, Ali Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *IJISAE*, vol. 7, pp. 88 - 91, 2019.
- [30] J. W. G. Putra, Pengenalan Konsep Pembelajaran Mesin dan Learning, Tokyo, 2020.
- [31] Shweta Kharya, Sunita Soni, "Weighted Naive Bayes Classifier: A Predictive Model for," *International Journal of Computer Applications*, vol. 133, pp. 32-37, 2016.
- [32] Palak Baid, Apoorva Gupta, Neelam Chaplot, "Sentiment Analysis of Movie Reviews using Machine," *International Journal of Computer Applications*, vol. 179, pp. 45-49, 2017.