

Analisis Sentimen Pengguna Media Sosial Terhadap Omnibus Law Berdasarkan Hashtag di Twitter

Analysis of Social Media Users Sentiments against Omnibus Law Based on Hashtags on Twitter

¹Okta Fanny*, ²Heri Suroyo

^{1,2}Sistem Informasi, Ilmu Komputer, Universitas Bina Darma,
Jl. Jendral Ahmad Yani No.3, 9/10 Ulu, Seberang, 9 Ulu, Seberang Ulu I,
Kota Palembang, Sumatera Selatan 30264, Indonesia.
e-mail: oktafannyz107@gmail.com

(received: 27 Oktober 2021, revised: 25 November 2021, accepted: 28 November 2021)

Abstrak

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter terhadap Omnibus Law. Setelah dilakukan analisis sentimen, terlihat sentimen netral dengan presentase terbesar yaitu 55%, selanjutnya sentiment positif sebesar 35% dan sentiment negative sebesar 10%. Hasil Analisis yang dilakukan memperlihatkan bahwa metode Naïve Bayes Classifier memberikan hasil pengujian klasifikasi dengan akurasi pada Hashtag Pro dengan nilai rata-rata akurasi mencapai 92.1%, nilai presisi dengan rata-rata 94.8% dan nilai recall dengan rata-rata 90.7%. Sedangkan Hashtag Kontra Untuk klasifikasi data, dengan nilai rata-rata akurasi mencapai 98.3%, nilai presisi dengan rata-rata 97.6% dan nilai recall dengan rata-rata 98.7%. Hasil dari Analisis text cloud yang dilakukan pada gabungan Hashtag baik Hashtag pro maupun Hashtag kontra, kata yang dominan muncul yaitu Omnibus Law yang artinya seluruh Hashtag yang di scrap memang benar mendiskusikan topik utama yaitu tentang Omnibus Law.

Kata kunci: Analisis Sentimen, Omnibus Law, Naïve Bayes Classifier

Abstract

Based on the research result, it can be concluded that Sentiment Analysis can be used to know the sentiment of the public, especially Twitter netizens against omnibus law. After the sentiment analysis, it looks neutral Sentiment with the largest percentage of 55%, then positive sentiment by 35% and negative sentiment by 10%. The results of the analysis showed that the Naïve Bayes Classifier method provides classification test results with accuracy in Hashtag Pro with an average accuracy score of 92.1%, precision values with an average of 94.8%, and recall values with an average of 90.7%. While Hashtag Counter For data classification, with an average accuracy value of 98.3%, precision value with an average of 97.6%, and recall value with an average of 98.7%. The result of text cloud analysis conducted on a combination of hashtags both Hashtag pros and Hashtags cons, the dominant word that appears is Omnibus Law which means that all hashtags in scrap are discussing the main topic that is about Omnibus Law.

Keywords: Sentiment Analysis, Omnibus Law, Naïve Bayes Classifier

1 Pendahuluan

Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, paragraf, dll) bersifat positif, *negative* atau netral. Analisis sentimen merupakan bidang penelitian yang cukup populer, karena dapat memberikan keuntungan untuk berbagai aspek, mulai dari prediksi penjualan[1]

Opini masyarakat dapat diperoleh dari berbagai media cetak maupun elektronik. Masyarakat pada saat ini lebih sering menggunakan media sosial dalam mengomentari suatu masalah termasuk suatu penomen. Salah satu media sosial yang digemari masyarakat Indonesia saat ini adalah *Twitter*. [2]

<http://sistemasi.ftik.unisi.ac.id>

Tidak sedikit masyarakat Indonesia menuangkan ulasan kritik atau pendapat mereka mengenai Omnibus Law ini melalui media sosial. Memang fitur komentar yang tersedia akan menimbulkan beragam perspektif baik positif maupun negatif, akan tetapi dengan banyaknya serta bervariasinya komentar yang netizen berikan tentu akan memakan banyak waktu dalam menganalisis pendapat netizen mengenai Omnibus Law ini. Adapun tagar terbanyak adalah #DemoBukanSolusi(11.632), #TolakOmnibusLaw (8.707) , sedangkan tagar yang mendukung Omnibus Law di antaranya #DukungOmnibusLaw (1.064), #ProOmnibusLaw (6696), #DemoBukanSolusi(11.632).

Dalam penelitian ini menggunakan pemrograman *Python*, *Python* adalah salah satu bahasa pemrograman tingkat tinggi yang bersifat *interpreter*, *interactive*, *object-oriented*, dan dapat beroperasi hampir di semua platform: *Mac*, *Linux*, dan *Windows*. *Python* termasuk bahasa pemrograman yang mudah dipelajari karena sintaks yang jelas, dapat dikombinasikan dengan penggunaan modul-modul siap pakai, dan struktur data tingkat tinggi yang efisien[3]

Tujuan dari penelitian ini adalah untuk menganalisis pada “Analisis sentimen pengguna media sosial terhadap *Omnibus Law* berdasarkan hastag di *Twitter* menggunakan pemrograman *Python*”. Melakukan Analisis sentimen pengguna media sosial terhadap *Omnibus Law* dari periode bulan Oktober sampai dengan bulan November 2020. Melakukan *data mining* dengan *web scrab Hashtag Omnibus Law* pada media sosial *Twitter* dengan membandingkan *Hashtag* yang pro dan yang kontra menggunakan pemrograman *Python*. Menerapkan penggunaan metode *Naïve Bayes Classifier* dengan *tools Rapid Miner* untuk menganalisis sentiment pengguna media sosial terhadap *Omnibus Law*. Manfaat dari penelitian ini adalah : Dapat menghitung persentase pendapat baik atau pendapat buruk tentang *Omnibus Law* agar dapat membandingkan hasil analisis sentimen. Dapat menganalisis sentimen pengguna media sosial terhadap *Omnibus Law* melalui metode *Naïve Bayes Classifier*. Dapat mengetahui kecenderungan sentimen berdasarkan *Hashtag* oleh Pengguna Media Sosial *twitter* mengenai *Omnibus Law*.

2 Tinjauan Literatur

2.1 Penelitian Sebelumnya

Dari penelitian yang telah dilakukan oleh (Buntoro 2017) yang berjudul, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen *Twitter* terhadap calon Gubernur DKI Jakarta 2017. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada *tweet* opini Bahasa Indonesia yang ada di *Twitter*. Setelah dilakukan analisis sentimen, terlihat berapa banyak sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017. Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi *Naïve Bayes Classifier* (NBC) untuk klasifikasi data AHY, dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Naïve Bayes Classifier* (NBC) lebih tinggi akurasinya untuk klasifikasi sentimen *Tweet* Bahasa Indonesia dibandingkan dengan metode klasifikasi *Support Vector Machine* (SVM). Untuk penelitian selanjutnya perlu dicoba dikembangkan dengan menggunakan data yang lebih banyak dan *Real Time*.

Penelitian yang kedua yang berjudul Analisis sentimen pasar otomotif mobil pada *tweet* *Twitter* dengan metode *Naïve Bayes* yang diteliti oleh Deden Rustiana pada Tahun 2017, Hasil penelitian ini menunjukkan bahwa tingkat akurasi *Naïve Bayes* yaitu 93%. Penelitian kedua yaitu sentimen analisis *tweet* berbahasa Indonesia dengan *Deep Belief Network* (DBN) dan hasil DBN dibandingkan dengan *Naïve Bayes* dan *Support Vector Machine* (SVM). Tujuan penelitian tersebut adalah untuk mengetahui hasil sentimen terhadap *tweet* berbahasa Indonesia di *Twitter* dan hasil akurasi DBN sebesar 93.31%, *Naïve Bayes* sebesar 79.10%, dan SVM sebesar 92.18%.

Berbeda dengan penelitian sebelumnya, tujuan dari penelitian ini adalah mengetahui hasil sentimen positif dan negatif terhadap keseluruhan data uji *tweet* pengguna media sosial *Twitter* yang menggunakan data sentiment per *hashtag* *Omnibus Law* dan dari hasil sentimen analisis tersebut dapat diinfokan ke publik untuk menggunakan media sosial terutama *Twitter* secara tepat. Selain itu juga, untuk mengetahui tingkat akurasi *Naïve Bayes Classifier*.

2.2 Analisis Sentimen

Analisis sentimen adalah Proses penentuan apakah isi suatu kumpulan data dalam bentuk teks (dokumen, kalimat, paragraf, dll) positif, negatif, atau netral. Analisis sentimen adalah bidang studi yang sangat populer karena dapat mengambil manfaat dari banyak aspek prakiraan penjualan [1]. Dalam arti lain, proses penentuan apakah sebuah kalimat afirmatif, negatif atau netral. Ada beberapa cara untuk melakukan analisis sentimen yang membutuhkan data. Misalnya, data opini pengguna dapat diperoleh dari media sosial dengan memeriksa tren media sosial dan mengonfirmasi ketersediaan berbagai data. Ada beberapa penamaan terhadap studi ini, yaitu analisis sentimen, penambangan opini (opinion mining), ekstraksi opini (opinion extraction), penambangan sentimen (sentiment mining) dimana semuanya sekarang berada dibawah ranah analisis sentimen atau penambangan opini. Dalam dunia industri kata “analisis sentimen” sering digunakan tetapi di dunia pendidikan kata “analisis sentimen” dan “opinion mining” sering digunakan. Kata “analisis sentimen” pertama muncul tahun 2003 oleh Nasukawa dan Yi, sementara “opinion mining” muncul pada tahun 2003 oleh Dave Lawrence dan Pennock[4]

2.3 Web Scrapping

Scraping Web (juga disebut panen *Web* atau *Web* ekstraksi data) adalah sebuah perangkat lunak komputer teknik penggalian informasi dari situs *web*. Biasanya, program perangkat lunak tersebut mensimulasikan eksplorasi manusia dari *Web* oleh salah satu rendah menerapkan-*Hypertext Transfer Protocol (HTTP)*, atau embedding *Web* browser tertentu penuh, seperti *Internet Explorer (IE)* dan *Mozilla Web browser*. *Web Scraping* berkaitan erat dengan pengindeksan *Web*, yang indeks konten *Web* menggunakan bot dan merupakan teknik universal yang diadopsi oleh kebanyakan mesin pencari. Sebaliknya, menggores *Web* lebih memfokuskan pada transformasi konten *Web* yang tidak terstruktur, biasanya dalam format. [5]

Scraping web dapat digunakan untuk berbagai skenario, seperti kerokan kontak, perubahan harga pemantauan/ perbandingan, pengumpulan ulasan produk, pengumpulan daftar real estat, cuaca pemantauan data, deteksi perubahan situs web, dan integrasi data web. Misalnya, pada skala mikro, harga saham dapat secara teratur dikerok untuk memvisualisasikan perubahan harga dari waktu ke waktu, dan umpan media sosial dapat secara kolektif dikerok untuk menyelidiki opini publik dan mengidentifikasi pemimpin opini. Pada tingkat makro, metadata hampir setiap situs web terus-menerus dikerok untuk membangun Internet mesin pencari, seperti Google Penelusuran atau Bing Pencarian. [6]

2.4 Naïve Bayes Classifier

Algoritma naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam penelitian ini yang menjadi data uji adalah dokumen tweets.[7]. *Naïve Bayes Classifier* merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi/kejadian. Sebelum menjelaskan *Naïve Bayes Classifier* ini, akan dijelaskan terlebih dahulu Teorema Bayes yang menjadi dasar dari metoda tersebut.

2.5 Text Mining

Text mining adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data. Ini juga dikenal sebagai analisis teks, meskipun beberapa orang menarik perbedaan antara dua istilah; dalam pandangan itu, analitik teks adalah aplikasi yang diaktifkan oleh penggunaan teknik *text mining* untuk memilah-milah set data.

Text mining telah menjadi lebih praktis bagi para ilmuwan data dan pengguna lain karena pengembangan platform data besar dan algoritma pembelajaran mendalam yang dapat menganalisis kumpulan data yang tidak terstruktur secara besar-besaran. Menganalisis teks membantu organisasi menemukan potensi wawasan bisnis yang berharga dalam dokumen perusahaan, email pelanggan, log call center, komentar survei *verbatim*, posting jaringan sosial, catatan medis dan sumber data berbasis teks lainnya. Semakin banyak, kemampuan penambangan teks juga dimasukkan ke dalam *AI chatbots*

dan agen virtual yang digunakan perusahaan untuk memberikan tanggapan otomatis kepada pelanggan sebagai bagian dari pemasaran, penjualan, dan operasi layanan pelanggan mereka [7].

3 Metode Penelitian

Waktu penelitian ini yaitu mengambil data tweet dari periode bulan Oktober sampai dengan bulan November, sedangkan proses Analisis data hasil dari web scraping data Twitter yaitu dari bulan Januari sampai dengan bulan Mei. pada masa tersebut peneliti melakukan tahapan demi tahapan sesuai pada perencanaan penelitian yang sebelumnya telah dilakukan terlebih dahulu.

3.1 Sumber Data

Data *tweet* dicari dan diambil dengan *Twitter API* pada *Twitter* dengan *keyword* “Omnibus Law” dan *Hashtag* yang pro dan *Hashtag* yang kontra, *Hashtag pro* antara lain yaitu, #RUUCiptakerLindungiPekerja, #TolakDemoDitengahPandemi, dan *Hashtag* yang kontra antara lain yaitu, #TolakOmnibusLaw, #GagalkanOmnibusLaw. Sistem menggunakan *user ID* dan *consumer key ID* pengguna pada *twitter* untuk dapat mengakses dan meretrieve *tweet* yang bersangkutan. Data yang didapat tersebut kemudian dipilah secara manual agar nantinya *tweet* yang dipakai murni berupa teks berbahasa Indonesia dan tidak mengandung gambar dan akan disimpan kedalam tabel *corpus_tweets*.

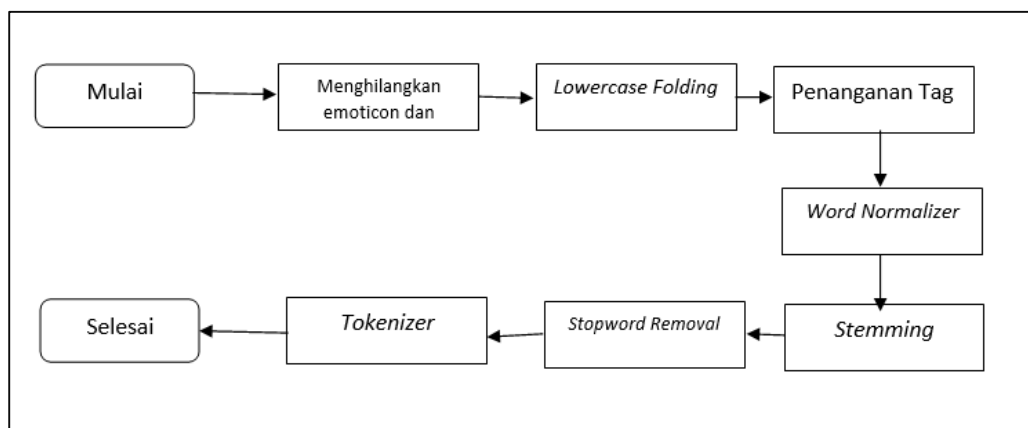
Dataset pada penelitian ini menggunakan format CSV yang dikumpulkan dari *Twitter* dengan metode *Web Scraping* dari media sosial *Twitter*. Data yang diambil hanya *tweet* dalam bahasa Indonesia, yaitu *tweet* dengan kata kunci #DemoBukanSolusi, #DukungOmnibusLaw, #ProOmnibusLaw untuk dataset *Hashtag* Pro Omnibus Law, dan #OmnibusLaw, #TolakOmnibusLaw untuk dataset *Hashtag* Kontra. Data diambil secara acak baik dari user biasa ataupun media *online* di *Twitter*.

3.2 Metode Pengumpulan Data

Web scraping adalah proses untuk mengekstraksi informasi dan data di dalam *website* secara otomatis. Data yang terekstrak tersebut kemudian disimpan ke dalam format yang diinginkan. Kita dapat menyimpan data tersebut dalam format teks, CSV, atau JSON. Dengan *web scraping*, seseorang dapat dengan mudah mengumpulkan data pelanggan dalam menentukan strategi pemasaran yang sesuai. *Web scraping* juga dapat kita gunakan untuk mengumpulkan data lain seperti keperluan untuk menganalisis data kompetitor. Karena tidak peduli jenis bisnisnya, kita akan selalu perlu melihat bagaimana kompetitor kita bekerja [9] *Web Scraping* terkait erat dengan *Web* pengindeksan, yang merupakan teknik pengambilan informasi yang diadopsi oleh beberapa mesin pencari untuk mengindeks informasi di *Web* melalui bot. Sebaliknya, *Web Scraping* berfokus pada transformasi data yang tidak terstruktur di *Web*, biasanya di Format HTML, ke dalam data terstruktur yang dapat disimpan dan dianalisis di *database* atau *spreadsheet* lokal pusat [10].

3.3 Pre-Processing Data

Pada tahap *pre-processing data*, data *tweet* mentah terlebih dahulu dilakukan proses *case folding*, *tokenizing*, *stemming*, serta *filtering*. Hasil dari tahapan ini menghasilkan fitur yang digunakan sebagai data pembelajaran mesin oleh NBC. *Natural Language Processing (NLP)* merupakan salah satu cabang ilmu Kecerdasan Buatan yang berfokus pada pengolahan Bahasa natural yang dapat digunakan untuk mengatasi hambatan pengenalan bahasa percakapan sehari hari oleh sistem komputer [11].



Gambar 1. Tahap Pra-Pemrosesan Menggunakan Pendekatan NLP

Alur kerja tahapan pra-pemrosesan dengan pendekatan NLP ditunjukkan pada **Gambar 1** , dengan penjelasan setiap tahapannya sebagai berikut:

1. *Menghilangkan emoticons dan symbol*
Pada penelitian ini emoticons dan simbol yang ada pada ulasan dibuang karena penelitian ini hanya berfokus pada teks yang terkandung dalam ulasan. Simbol yang dibuang adalah “~”, “\”, “!”, “\$”, “%”, “^”, “&”, “*”, “(”, “)”, “_”, “-”, “+”, “=”, “:”, “;”, “”, “>”, “<”, “>”, “koma”, “titik”, “?”, “/”, “\”, “#”, dan “|”. Ulasan yang mengandung simbol seperti “ayo kita bersama gagalkan omnibus Law!!!” setelah diproses maka menjadi “ayo kita bersama gagalkan omnibus Law”
2. *Lowercase folding*
Lowercase folding adalah mengubah semua huruf menjadi huruf kecil agar terhitung sebagai kata yang sama. Komentar “Ayo kita cegat pengesahan RUU CIPTAKER” menjadi “ayo kita cegat pengesahan ruu ciptaker”, huruf “A, RUU CIPTAKER” kapital diubah menjadi “a, ruu ciptaker” huruf kecil.
3. *Penanganan Hashtag*
Hashtag adalah potongan informasi yang telah disediakan *Twitter* dan dapat disertakan isi *twit* pengguna *Twitter*. Tidak semua *Twit* mengandung *Hashtag*, sehingga perlu dilakukan penanganan khusus pada *twit* yang mempunyai tag. *Hashtag* pada *twit* ditambahkan menjadi bagian dari isi komentar ulasan. Contoh *Hashtag* yang terdapat pada *twitter* seperti “#GagalkanOmnibusLaw”, “GagalkanOmnibusLaw”.
4. *Word Normalizer*
Word Normalizer digunakan untuk memperbaiki kata-kata dalam ulasan sehingga menghasilkan kalimat yang baik dan benar sesuai dengan aturan tata bahasa Indonesia. Peningkatan ini diperlukan untuk memudahkan pembaca memahami makna kalimat. Komentar “kerumunan pedemo RUU Ciptaker ini padat bgt” maka setelah proses *Word Normalizer* menjadi “kerumunan pendemo RUU Ciptaker ini padat banget”. Kata “bgt” diubah menjadi “banget” sehingga lebih mudah dipahami.
5. *Stemming*
Stemming berfungsi untuk membuat suatu kata menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada kata tersebut. Sebagai contoh kalimat “Ayo Kita Cegat Pengesahan RUU CIPTAKER” kemudian diubah menjadi “Ayo kita cegat sah RUU CIPTAKER”.
6. *Stopword Removal*
Stopword removal berfungsi untuk menghilangkan kata-kata yang mempunyai jumlah kemunculan banyak tapi tidak terlalu penting. Kata yang masuk *stopword* seperti “yang”, “dan”, “di”, “dari” sehingga menyisakan kata-kata yang penting.
7. *Tokenizer*
Tokenizer berfungsi untuk membagi teks input menjadi *array token*, karakter yang digunakan dalam bahasa Indonesia adalah alfabet dan setiap token dipisahkan oleh spasi.

Sebagai contoh kalimat “Ayo kita cegat pengesahan RUU CIPTAKER” kemudian diubah menjadi “Ayo,” kita”, “cegat”, “pengesahan”, “RUU” “CIPTAKER”

3.4 Labeling Data

Pada konteks analisis sentimen berbahasa Indonesia, sebenarnya sudah ada dataset-dataset yang dipublikasikan secara umum, misalnya dataset dari penelitian tweet bahasa Indonesia nonformal yang berisi 4.000 tweet dengan label polaritas utama, yaitu positif, negatif, dan netral, atau Indonesian-Emotion-Twitter-Dataset, yang berisi 4.403 tweet dengan polaritas emosi yang lebih spesifik (love, joy, anger, sadness, fear) [12].

Namun, dataset berbahasa Indonesia yang lebih banyak ditemukan adalah data yang berasal dari penelitian-penelitian analisis sentimen yang memiliki cakupan yang lebih spesifik dan kurang cocok untuk digunakan pada analisis sentimen yang jangkauannya lebih luas, misalnya analisis sentimen terhadap operator seluler Indonesia yang berjumlah 1.000 tweet yang berisi komentar mengenai operator seluler dan dibagi menjadi dua polaritas, yaitu positif dan negatif. Beberapa contoh lainnya yaitu analisis sentimen yang berfokus pada [13].

Tweet yang sudah diseleksi ini kemudian diberi label secara manual menggunakan tiga jenis polaritas, yaitu polaritas positif (1) untuk melambangkan tweet yang memiliki sentimen yang bersifat positif, misalnya persetujuan dan kebahagiaan; polaritas negatif (-1) untuk melambangkan tweet yang memiliki sentimen negatif, seperti penolakan atau amarah dan kekecewaan; dan yang terakhir polaritas netral (0) untuk tweet yang tidak menunjukkan sentimen positif ataupun negatif. Distribusi label positif, negatif, dan netral pada *Indonesian-General-Sentiment-Analysis-Dataset* [14]. Dari segi *pre-processing*, pada *dataset* ini telah dilakukan pembersihan elemen-elemen pengganggu pada data, misalnya simbol-simbol, tanda baca, angka, tautan halaman *web*, *Hashtag*, dan *mention*. Pembersihan data ini dilakukan guna meningkatkan kualitas data saat digunakan untuk pelatihan model analisis sentimen. Selain itu, juga disediakan beberapa versi data dengan penerapan teknik *pre-processing* tambahan, yaitu *stemming*, yang mengubah kata menjadi bentuk dasarnya dan penghilangan kata sambung [13]. Seperti terlihat pada Tabel 1

Tabel 1. Sentiment Data Twitter Dengan Keyword “ Omnibus Law”

Sentiment	Tweet
-1	aksi demo besar buruh senin november puluh ribu massa pusat mahkamah konstitusi
-1	aksi tolak omnibuslaw kantor gubernur jatim umur juang
0	aktor balik provokator demo tolak omnibuslaw ya
1	alhamdulillah manfaat omnibuslaw
-1	bahaya tular omnibuslaw
0	bentar td ente bilang gak ngaitkan ante dg prabowo kaitka
-1	blacklist aja demo dunia kerja biar manfaat omnibuslaw stop nyari sakit
1	demo bawa sakit omnibuslaw

3.5 Analisis Sentimen

Setelah menganalisis text cloud dan memberikan gambaran jumlah data yang akan dikelola , proses selanjutnya yaitu Analisis sentiment dari data yang telah terkumpul yang berjumlah 2667 data tweet yang mana jumlah data *Hashtagh* pro nya berjumlah 1401 tweet dan data *Hashtag negative* berjumlah 1221 tweet. Di proses Analisis sentiment ini akan berisikan hasil dari perhitungan dari jumlah sentiment pengguna media sosial terhadap Omnibuslaw yang selanjutnya akan di olah lagi menggunakan algoritma Naïve Bayes.

4 Hasil dan Pembahasan

4.1 Hasil Data Pre-Processing

Text asli sebelum dan sesudah di bersihkan menggunakan metode *pre-processing* dapat dilihat pada Gambar 2.

text	remove_user
RT @Robert_Moses2: Stop Nyari Penyakit\n#DemoC...	RT : Stop Nyari Penyakit\n#DemoCiptaClusterKov...
Jenis-jenis Omnibuslaw\nDemo Nyusahin Rakyat #...	Jenis-jenis Omnibuslaw\nDemo Nyusahin Rakyat #...
ga usah demo\n#omnibuslaw https://t.co/yTDKTLZFIN	ga usah demo\n#omnibuslaw https://t.co/yTDKTLZFIN
yang ada bakal bencana\n#omnibuslaw https://t...	yang ada bakal bencana\n#omnibuslaw https://t...
Omnibuslaw meningkatkan kinerja Investasi\n#ND...	Omnibuslaw meningkatkan kinerja Investasi\n#ND...
...	...
Kita harus saling menjaga agar terhindar dari ...	Kita harus saling menjaga agar terhindar dari ...
RT @Lini_ZQ: Siang ini, 3 mahasiswa melakukan ...	RT : Siang ini, 3 mahasiswa melakukan aksi sim...
Kita harus saling menjaga agar terhindar dari ...	Kita harus saling menjaga agar terhindar dari ...
@tempodotco Wakil rakyat dalam pengesahan peng...	Wakil rakyat dalam pengesahan pengebutan uu o...
Kalau begini... \nFixed. Jomblo makin banyak ...	Kalau begini... \nFixed. Jomblo makin banyak ...

Gambar 2 Hasil dari metode *Pre-Processing*

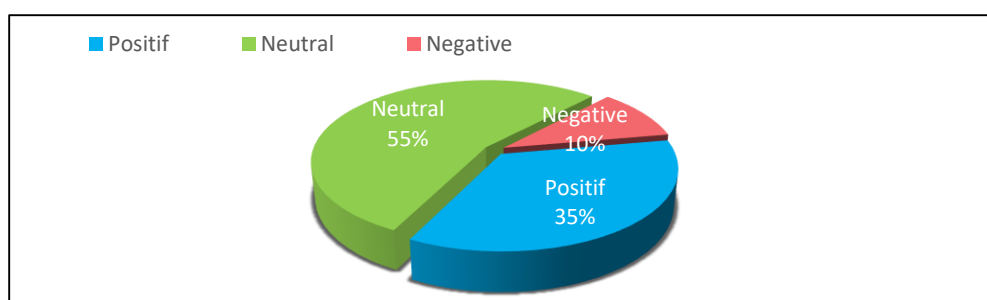
4.2 Hasil Analisis

Dari lima *Hashtag* yang sudah dilakukan Analisis sentiment dapat disimpulkan bahwa keberagaman pendapat pengguna media sosial terhadap topik Omnibuslaw ini, hal ini dapat dilihat dari rincian pada Tabel 2 Berikut :

Tabel 2. Rincian Hasil Keseluruhan Data Sentiment Keseluruhan Hashtag

Hashtag	Sentiment						Jumlah
	positif	%	Neutral	%	Negative	%	
#DukungOmnibusLaw	187	41,2 %	202	44,5 %	64	14,1 %	453
#DemoBukanSolusi	181	37,4 %	219	45,3 %	83	17,1 %	483
#ProOmnibusLaw	216	42,3 %	217	42,5 %	77	15,1 %	510
#OmnibusLaw	93	16,9 %	234	42,7 %	221	40,3 %	548
#TolakOmnibusLaw	95	14,1 %	329	48,8 %	249	37,0 %	673
TOTAL	772	35%	1202	55%	694	10%	2667

Berdasarkan Tabel 2 di atas dapat di lihat bahwa sentiment *Neutral* dengan persentasi terbesar yaitu 55% selanjutnya sentiment positif dengan persentasi 35% dan sentiment Negatif 10%, besarnya persentasi sentiment neutral hal ini dikarenakan banyaknya data yang bersifat menengahi dan tidak terlalu ikut berdebat mengenai apakah Omnibuslaw ini baik atau tidak, mereka malah mencemaskan kondisi pendemo yang berkerumun ditengah pandemic, selain itu banyaknya *Tweet* yang tidak menggunakan kata kasar ataupun kata yang bersifat positif yang menyebabkan proses labeling tidak menganggap data tersebut masuk ke dalam kategori data yang bersifat *negative* atau pun positif

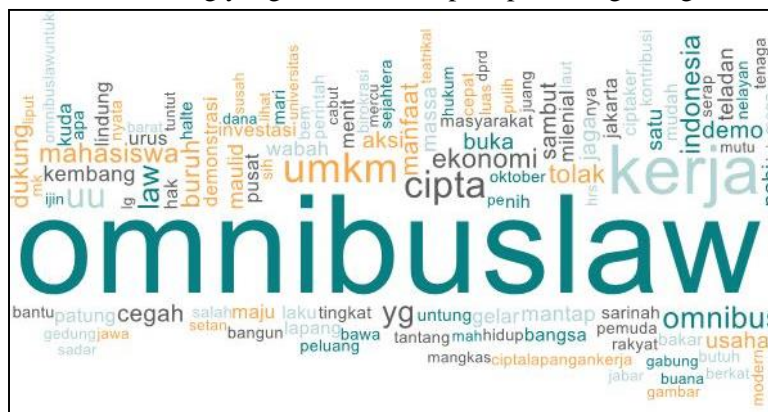


Gambar 3. Diagram Persentase Sentiment keseluruhan data

Berdasarkan diagram diatas dapat di lihat bahwa sentiment Neutral dengan persentasi terbesar yaitu 55% selanjutnya sentiment positif dengan persentasi 35% dan sentiment Negatif 10%, besarnya persentasi sentiment neutral hal ini dikarenakan banyaknya data yang bersifat menengahi dan tidak

terlalu ikut berdebat mengenai apakah Omnibuslaw ini baik atau tidak, mereka malah mencemaskan kondisi pendemo yang berkerumun ditengah pandemic, selain itu banyaknya Tweet yang tidak menggunakan kata kasar ataupun kata yang bersifat positif yang menyebabkan proses labeling tidak menganggap data tersebut masuk ke dalam kategori data yang bersifat negative atau pun positif.

Ada total 5 Hashtag dalam data sistem dan jumlah data tweet yang diterapkan adalah 2.667 tweet, 483 tweet #DemoBukanSolusi, 453 tweet #DukungOmnibuslaw, 548 tweet #Omnibuslaw, 510 tweet #ProOmnibuslaw, 673 tweet #TolakOmnibuslaw diambil dari 170 data dari setiap Hashtag ketika data pengujian mencapai 850. Jumlah pengguna twitter yang tersaring yang ikut serta dalam tweet dengan Hashtag pro dan kontra ini berjumlah 1481 Pengguna, data yang berjumlah 2667 tersebut adalah jumlah data bersih yang sudah di kelolah dalam program pre-processing Dapat dilihat dari Gambar 4 bahwa hasil dari Text Cloud dari kedua Hashtag yang sudah terkumpul apabila digabungkan.



Gambar 4. Hasil Text Cloud dari kedua hashtag pro dan kontra

Dapat dilihat dari cloud text diatas, kata yang dominan muncul yaitu Omnibus Law yang artinya seluruh Hashtag yang di scrap memang benar mendiskusikan topik utama yaitu tentang Omnibus Law. tapi disamping membahas omnibuslaw, pengguna sosial media juga menyinggu beberapa hal lain seperti UMKM, Cipta Kerja, dan juga ekonomi yang artinya kedua belah kelompok sepakat bahwa fokus permasalahan utama diskusinya ada pada kata tersebut. Jika diperhatikan lebih detail bahkan muncul kata milenial, ini harusnya menarik dibahas, karena itu artinya yang dominan melakukan diskusi ini bisa jadi adalah kelompok usia milenial. Fenomena Omnibuslaw ternyata menarik perhatian dari kalangan milenial untuk ikut serta dalam diskusi mengenai Omnibuslaw ini. Di setiap tweet yang tersaring ke dalam data scrap pada penelitian ini semuanya mengandung kata Omnibuslaw, sehingga data ini bisa dikatakan bahwa data yang valid untuk diteliti karena sesuai dengan topik utama pembahasan ini.

Untuk mengetahui akurasi, Analisis Sentimen pengguna Media sosial Twitter terhadap Omnibus Law diklasifikasi menggunakan metode *Naïve Bayes Classifier (NBC)* dengan perangkat lunak *Rapid Miner* versi 9.8.001 menggunakan tipe dokumen Comma Separated Values (CSV) sebagai inputan data untuk melakukan klasifikasi data.

Perbandingan hasil dari metode klasifikasi *Naïve Bayes Classifier (NBC)* terhadap *Hashtag Pro* dan *Hashtag Kontra* dapat dilihat pada tabel 3 Berikut

Tabel 3. Hasil Klasifikasi Menggunakan Naive Bayes

<i>Hashtag Pro</i>			
<i>Sentiment</i>	<i>Recall (%)</i>	<i>Precision (%)</i>	<i>Accuracy (%)</i>
<i>Positif</i>	87.8	100	92.1%
<i>Neutral</i>	100.0	85.9	
<i>Negative</i>	84.5	98.7	
<i>Hashtag Kontra</i>			
<i>Sentiment</i>	<i>Recall (%)</i>	<i>Precision (%)</i>	<i>Accuracy (%)</i>
<i>Positif</i>	92.2	100	98.3%
<i>Neutral</i>	100	97.3	
<i>Negative</i>	99.1	98.9	

Tabel 3 berisi informasi mengenai nilai akurasi, presisi dan *recall* dari masing-masing uji coba yang telah dilakukan. Bagian kolom berisi informasi mengenai Sentiment. Sedang bagian baris berisi nilai akurasi, presisi dan *recall* dari masing-masing uji coba yang telah dilakukan. Dari proses data *preprocessing* menghasilkan sejumlah token yang kemudian digunakan sebagai input sebuah proses klasifikasi. Proses klasifikasi dilakukan menggunakan metode *Naïve Bayes Classifier (NBC)*. Dari proses klasifikasi diperoleh nilai akurasi, presisi dan *recall* dari masing-masing uji coba.

Dari penelitian yang telah dilakukan, diketahui bahwa hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi *Naïve Bayes Classifier (NBC)* untuk klasifikasi data *Hashtag Kontra*, sedangkan untuk nilai akurasi terendah didapat saat metode klasifikasi *Naïve Bayes Classifier (NBC)* digunakan untuk klasifikasi data *Hashtag Pro*. Meskipun menghasilkan akurasi yang cukup baik, tapi model yang dibangun masih melakukan sedikit kesalahan pada saat proses klasifikasi data yang pembagian sentimennya tidak seimbang. Karena dengan menggunakan data yang tidak seimbang akan menyebabkan data *minority class* yang salah diklasifikasi sebagai data *majority class*[15], pada akhirnya menjadikan selisih nilai menjadi besar.

5 Kesimpulan

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen *Twitter* terhadap *Omnibus Law*. Setelah dilakukan analisis sentimen, terlihat sentimen netral dengan presentase terbesar yaitu 55%, selanjutnya sentimen positif sebesar 35% dan sentiment negative sebesar 10 %.

Hasil Analisis yang dilakukan memperlihatkan bahwa metode *Naïve Bayes Classifier* memberikan hasil pengujian klasifikasi dengan akurasi pada *Hashtag Pro* dengan nilai rata-rata akurasi mencapai 92.1%, nilai presisi dengan rata-rata 94.8% dan nilai *recall* dengan rata-rata 90.7%. Sedangkan *Hashtag Kontra* Untuk klasifikasi data, dengan nilai rata-rata akurasi mencapai 98.3%, nilai presisi dengan rata-rata 97.6% dan nilai *recall* dengan rata-rata 98.7%.

Hasil dari Analisis *text cloud* yang dilakukan pada gabungan *Hashtag* baik *Hashtag pro* maupun *Hashtag kontra*, kata yang dominan muncul yaitu *Omnibus Law* dengan jumlah kata 1830 kata yang artinya seluruh *Hashtag* yang di *scrap* memang benar mendiskusikan topik utama yaitu tentang *Omnibus Law*.

Referensi

- [1] V. Chandani, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.
- [2] N. Anggraini and H. Suroyo, "Comparison of Sentiment Analysis against Digital Payment 'T-cash and Go-pay' in Social Media Using Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 1, no. 2, pp. 152–163, 2019, doi: 10.33557/journalisi.v1i2.21.
- [3] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [4] Liu B, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, pp. 1–167, 2012.
- [5] M. S. Utomo, "Web Scraping pada Situs Wikipedia menggunakan Metode Ekspresi Regular," *J. Teknol. Inf. Din.*, vol. 18, no. 2, pp. 153–160, 2013.
- [6] B. Zhao, "Encyclopedia of Big Data," *Encycl. Big Data*, no. December, 2020, doi: 10.1007/978-3-319-32001-4.
- [7] D. Rustiana and N. Rahayu, "Analisis sentimen pasar otomotif mobil:," *J. SIMETRIS*, vol. 8, no. 1, pp. 113–120, 2017.
- [8] T. Kwartler, "What is Text Mining?," *Text Min. Pract. with R*, pp. 1–15, 2017, doi: 10.1002/9781119282105.ch1.
- [9] 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, doi: 10.22146/ijccs.24716.
- [10] E. Vargiu and M. Urru, "Exploiting web scraping in a collaborative filtering- based approach to <http://sistemasi.ftik.unisi.ac.id>

- web advertising,” *Artif. Intell. Res.*, vol. 2, no. 1, pp. 44–54, 2012, doi: 10.5430/air.v2n1p44.
- [11] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [12] S. Suryono, E. Utami, and E. T. Luthfi, “Klasifikasi Sentimen Pada Twitter Dengan Naive Bayes Classifier,” *Angkasa J. Ilm. Bid. Teknol.*, vol. 10, no. 1, p. 89, 2018, doi: 10.28989/angkasa.v10i1.218.
- [13] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, “Dataset Indonesia untuk Analisis Sentimen,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 334, 2019, doi: 10.22146/jnteti.v8i4.533.
- [14] F. T. Industri, “Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing,” *J. Inform.*, vol. 2, no. 1, pp. 13-22–22, 2001, doi: 10.9744/informatika.2.1.pp.13-22.
- [15] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.