

ANALISIS SENTIMEN MEDIA SOSIAL TWITTER TERHADAP REAKSI MASYARAKAT PADA RUU CIPTA KERJA MENGGUNAKAN METODE KLASIFIKASI ALGORITMA NAIVE BAYES

ANALYSIS OF TWITTER SOCIAL MEDIA SENTIMENT ON THE PUBLIC'S REACTION TO THE DRAFTS OF JOB CREATION LAW USING THE CLASSIFICATION METHOD NAIVE BAYES ALGORITHM

Delima Ayu Wulandari¹, Rd. Rohmat Saedudin², Rachmadita Andreswari³

^{1,2,3} Universitas Telkom, Bandung

delimaayu@student.telkomuniversity.ac.id¹, rdrohmat@telkomuniversity.ac.id², andreswari@telkomuniversity.ac.id³

ABSTRAK

Perkembangan teknologi pada zaman sekarang memberikan kemudahan bagi masyarakat dalam menggunakan media sosial sebagai salah satu sarana untuk mengekspresikan pikiran, pendapat hingga dengan melakukan kampanye atau demo. Salah satu media sosial yang masih banyak digunakan oleh masyarakat adalah Twitter. Studi kasus yang diambil oleh peneliti adalah kasus RUU Cipta Kerja. Kasus tersebut diambil karena RUU Cipta Kerja ramai diperbincangkan masyarakat Indonesia di media sosial Twitter. Salah satu pemanfaatan penelitian ini adalah untuk mengetahui kecenderungan komentar atau *tweets* pengguna Twitter terhadap adanya RUU Cipta Kerja dengan melakukan analisis sentimen. Sentimen akan diklasifikasikan dengan label positif, negatif dan netral. Algoritma yang digunakan dalam melakukan analisis sentimen adalah *Naive Bayes* karena memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan analisis sentimen. Tahapan dalam melakukan analisis sentimen pada penelitian ini adalah *preprocessing* data, pengolahan data, klasifikasi, dan evaluasi. Hasil dari Penelitian ini yaitu menggunakan rasio *training* dan *testing* yaitu 75:25 dikarenakan memiliki akurasi yang tertinggi yaitu 88% dibandingkan dengan rasio lainnya. Hasil prediksi yang dihasilkan dari klasifikasi ternyata lebih dominan kepada label positif. Selain melakukan klasifikasi untuk analisis sentimen, penelitian ini juga mengukur performa dari model yang dibuat. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki nilai *precision* 92%, *recall* 84%, dan *f1-score* 86%.

Kata kunci: *Twitter, Naive Bayes, RUU Cipta Kerja, Analisis Sentimen*

ABSTRACT

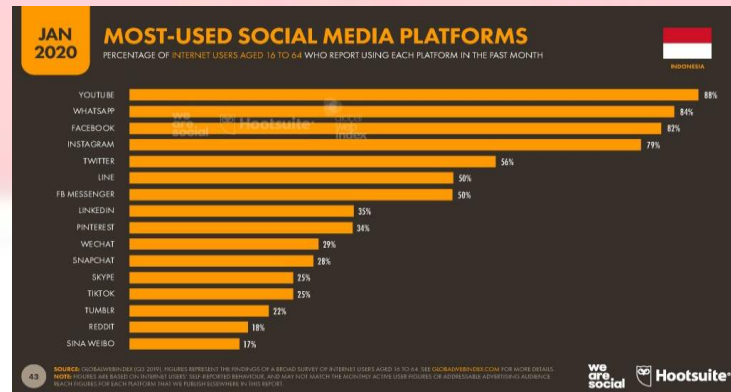
The development of the current technological era makes it easy for people to use social media as a means to convey thoughts, opinions to campaigns or demos. One of the social media that is still widely used by the public is Twitter. The case study takes by the researcher is a case of the Drafts of Job Creation Law. The case takes because the campaign for the Rejection of the Drafts of Job Creation Law widely discussed by the Indonesian people on Twitter social media. One of the uses of this research is to determine the tendency of Twitter users' comments or tweets to the existence of the Job Creation Bill by conducting sentiment analysis. Sentiment will be the classification by positive, negative, and neutral labels. The algorithm used in conducting sentiment analysis is Naive Bayes because it has a high degree of accuracy in classification sentiment analysis. The stages in conducting a sentiment analysis in this study are data preprocessing, data processing, classification, and evaluation. In the results of this study, researchers use the training and testing ratio of 75:25 because it has the highest accuracy of 88% compared to other ratios. The prediction results generated from the classification turned out to be more dominant to the positive label. In addition to classifying for sentiment analysis, this study also measures the performance of the model created. The results showed that the Naive Bayes algorithm has a precision value of 92%, recall 84%, and f1-score 86%.

Keyword: *Twitter, Naive Bayes, Drafts of Job Creation Law, Sentiment Analysis*

1. Pendahuluan

Perkembangan teknologi yang terjadi di Indonesia saat ini, sangat memudahkan pengguna atau masyarakat dalam mengakses dan menerima informasi itu baik melalui media sosial, media cetak, media *digital* maupun media sosial. Dengan berbagai informasi, masyarakat dengan bebas dapat mengeluarkan pendapatnya dan menyebarkan informasi yang didapat melalui media sosial yang dimiliki. Media sosial merupakan sebuah *platform* media yang digunakan untuk melakukan komunikasi satu sama lain yang dilakukan secara online sehingga memungkinkan seseorang dapat berinteraksi tanpa dibatasi waktu dan ruang.

Berdasarkan data dari wearesocial.com dan Hootsuite, perusahaan analisis media sosial yang berasal dari Kanada bahwa penggunaan media sosial aktif di Indonesia pada tahun 2020 sudah mencapai 59% atau 160 juta jiwa dari jumlah total penduduk Indonesia yang mencapai 272,1 juta jiwa dan pengguna internet sudah mencapai 64% atau 175,5 juta jiwa. Salah satu contoh media sosial di Indonesia yang paling banyak digunakan sampai saat ini adalah *Twitter*. Dapat dilihat pada Gambar 1.1 menurut wearesocial.com bahwa tingkatan media sosial di Indonesia per Januari 2020 yang paling banyak digunakan adalah *Youtube* sebanyak 88%, kemudian di urutan kedua *Whatsapp* 84%, urutan ketiga dengan jumlah pengguna 82% dari total penggunaan media sosial adalah *Facebook*. Urutan berikutnya adalah platform berbagai gambar *Instagram* 79% dan disusul oleh *Twitter* pada urutan kelima sebanyak 56%.



Gambar 1. Tingkatan Media Sosial di Indonesia

Dengan penggunaan media sosial *Twitter* saat ini, pengguna dapat mengeluarkan pendapatnya terhadap isu yang sedang marak dibicarakan. Contohnya saat ini, dimana sedang marak dibicarakan terkait penolakan Rancangan Undang-Undang (RUU) Cipta Kerja, sehingga banyak membuat para mahasiswa dari berbagai universitas di Indonesia dan buruh melakukan aksi demonstrasi. Demonstrasi tersebut dilakukan sebagai salah satu bentuk aksi kritikan terhadap pemerintahan Indonesia yang selalu terburu-buru dalam proses pembuatan dan pengesahan dalam berbagai kebijakan di Indonesia. Berdasarkan berita yang terkait dengan adanya demonstrasi mengenai penolakan terhadap RUU Cipta Kerja yang terjadi baru ini, pemanfaatan media sosial banyak digunakan untuk mengetahui banyaknya komentar ataupun pendapat masyarakat. Salah satu media sosial yang bisa melihat komentar serta trending pembicaraan yang lagi banyak dibicarakan yaitu *Twitter*. Analisis sentimen ini dapat digunakan untuk menyaring komentar-komentar di media sosial yang terkait dengan topik yang diinginkan baik itu pendapat positif serta negatif yang bisa berisikan informasi-informasi berguna yang menjadi fokus pada penelitian ini.

Dalam melakukan analisis sentimen ini berguna untuk mengetahui komentar masyarakat Indonesia terhadap adanya Rancangan Undang-Undang (RUU) Cipta Kerja pada situs media sosial *Twitter*, sehingga dibutuhkan adanya pengklasifikasian dengan adanya metode penunjang terhadap komentar ataupun tanggapan berupa teks yaitu *text mining*. Metode *text mining* berguna dalam menganalisis, pendapat, evaluasi, sikap, penilaian, emosi seseorang sehingga dapat diketahui apakah berkenaan dengan suatu topik, layanan, organisasi maupun kegiatan tertentu [1]. Setelah mengetahui dan mengklasifikasikan analisis dari komentar-komentar melalui *Twitter* terhadap isu RUU Cipta Kerja, penulis melakukan analisis data menggunakan klasifikasi *Naïve Bayes*.

Berdasarkan penelitian Deni Gunawan, Dwiza Riana, Dian Ardiansyah, Fajar Akbar, dan Salman Alfarizi pada tahun 2020 tentang analisis sentimen calon gubernur Jabar 2018-2023 menggunakan *Support Vector Machine* dan *Naïve Bayes* sebagai algoritma klasifikasinya [2]. Penelitian tersebut menyimpulkan bahwa *Naïve Bayes* adalah *classifier* terbaik dengan akurasi 93,29%, melebihi akurasi *Support Vector Machine* yaitu 92,61%. Berdasarkan penelitian-penelitian yang ada menunjukkan bahwa algoritma *Naïve Bayes* mempunyai tingkat akurasi yang lebih tinggi. Sehingga penelitian ini penulis menggunakan algoritma *Naïve Bayes* untuk melakukan analisis media sosial *Twitter* terhadap reaksi masyarakat pada RUU Cipta Kerja.

2. Dasar Teori

2.1 RUU Cipta Kerja

Rancangan Undang-Undang (RUU) Cipta Kerja atau yang sering disebut dengan *Omnibus Law* adalah sebuah rancangan metode untuk membuat suatu regulasi atau undang-undang yang terdiri atas banyaknya subjek atau materi pokok tertentu yang bertujuan agar menyimpangi suatu norma peraturan [3]. Rancangan undang-undang ini dirancang untuk memangkas sebagian norma yang dikira tidak cocok dengan

perkembangan zaman serta merugikan kepentingan bagi negeri, sehingga pemerintah memandang kalau RUU Cipta Kerja ini perlu dicoba karena tingginya angka pengangguran di Indonesia yang telah mencapai 7 juta jiwa serta diharapkan RUU Cipta Kerja dapat membuka lapangan kerja baru [4].

2.2 Twitter

Twitter adalah sebuah platform media sosial yang mempunyai layanan *microblogging* yang dimana mengizinkan penggunanya untuk mengirim pesan secara *real-time* [5]. Pesan di *Twitter* terkenal dengan nama *tweet*. Para pengguna *Twitter* dapat mengirimkan 140 karakter untuk dipost di *Twitter* tersebut, dari 140 karakter ini bisa berupa pujian ataupun komentar pada suatu hal yang dibahas dan juga karena keterbatasan karakter yaitu 140 yang bisa dituliskan, sehingga *tweet* seringkali mengandung bahasa singkatan, bahasa *slang* maupun kesalahan pengerjaan [6]. *Twitter* juga mampu menjadi wadah atau tempat berbagi ide ataupun informasi yang singkat dengan jumlah yang tidak terbatas, sehingga dari berbagai informasi inilah *Twitter* menjadi sumber yang digunakan pada *text classification*.

2.3 Algoritma Naïve Bayes

Naive Bayes Classifier adalah salah satu algoritma pada data mining yang termasuk kedalam metode klasifikasi. *Naive Bayes Classifier* merupakan metode klasifikasi yang menggunakan perhitungan probabilitas dan statistika yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes* [7]. Berikut adalah rumus *Naive Bayes* [8]:

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

Keterangan:

X	= Data dengan <i>class</i> yang belum diketahui
H	= Hipotesis data x merupakan suatu class spesifik
P(H X)	= Probabilitas hipotesis H berdasarkan kondisi x
P(H)	= Probabilitas hipotesis H
P(X H)	= Probabilitas X berdasarkan kondisi tersebut
P(X)	= Probabilitas dari X

2.4 Preprocessing

Preprocessing adalah langkah terpenting dalam melakukan analisis sentimen. *Preprocessing* atau pemrosesan text merupakan mempersiapkan sebuah teks menjadi daya yang siap diolah lebih lanjut dan bertujuan untuk menghilangkan *noise*, serta mengambil fitur penting yang ada pada dokumen teks [9]. *Preprocessing* dalam penelitian ini terdiri dari enam tahap yaitu *case folding*, *punctuation removal*, *tokenizing*, *normalization*, *stopword removal* dan *stemming*

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan suatu metode yang menentukan seberapa jauh keterhubungan kata terhadap dokumen dengan memberikan bobot tiap kata. TF-IDF merupakan gabungan dari konsep *Term Frequency (TF)* serta *Inverse Document Frequency (IDF)* yang menghitung frekuensi kemunculan suatu kata dalam dokumen serta *inverse* frekuensi dokumen yang memiliki kata tersebut [10].

$$TF(t,d) = \frac{\text{Number of times } t \text{ appear in document } d}{\text{Total number of terms in document } d}$$

$$IDF(t,d) = \log \left(\frac{\text{Total number of documents } D}{\text{Numbers of documents with the term in it}} \right)$$

$$TFIDF(t,d,D) = TF(t,d) * IDF(t,d)$$

2.6 Analisis Sentimen

Analisis sentimen adalah sesuatu yang menganalisa pendapat, sentimen, evaluasi, dan emosi orang tentang suatu topik yang yang diekspresikan melalui teks [11]. Tujuan analisis sentimen adalah menemukan pendapat atau ulasan, berdasarkan pendapat orang yang diklasifikasikan pada berbagai polaritas seperti positif, negatif, atau netral. Analisis sentimen juga dapat menyatakan perasaan seperti emosional sedih, gembira atau marah [12].

2.7 Evaluasi Performansi

Evaluasi Performansi adalah suatu tahapan untuk mengukur performansi dari *classifier* dalam melakukan klasifikasi yang digunakan dalam penelitian [13]. Dalam membuat evaluasi performansi matrik perlu adanya *confusion matrix*, yaitu suatu alat ukur berupa matriks yang digunakan buat memperoleh jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai [14].

Tabel 1. Struktur *Confusion Matrix*

	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

Jika model melakukan prediksi terhadap komentar dan menebak label positif dengan benar, maka pada Tabel 1 dapat dihitung sebagai *True Positive* (TP). Begitu juga dengan label negatif, jika ditebak dengan benar maka dapat dihitung sebagai *True Negative* (TN). Sebaliknya, jika model melakukan prediksi terhadap komentar namun menebak label positif dengan salah maka dapat dihitung sebagai *False Positive* (FP). *False Negative* (FN) juga merupakan hasil prediksi yang salah terhadap label negatif. Dengan adanya *Confusion Matrix*, performa sebuah model akan dihitung dari seberapa banyak model tersebut melakukan prediksi dengan benar.

2.7.1 Precision

Precision adalah tingkat ketepatan suatu informasi sistem dalam memprediksi target positif terhadap jumlah data yang diprediksi positif [15]. Berikut adalah rumus untuk menghitung presisi:

$$Precision = \frac{TP}{TP + FP}$$

2.7.2 Recall

Recall adalah tingkat ketepatan suatu informasi sistem dalam memprediksi target positif terhadap jumlah data yang riil positif [15]. Berikut adalah rumus untuk menghitung *recall*:

$$Recall = \frac{TP}{TP + FN}$$

2.7.3 F1-Score

F1-Score adalah salah satu perolehan dari konstanta dua dikali nilai *recall* dan *presisi*, dan dibagi jumlah keduanya [15]. Rumus F1-Score yaitu sebagai berikut:

$$F1-Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

2.7.4 Accuracy

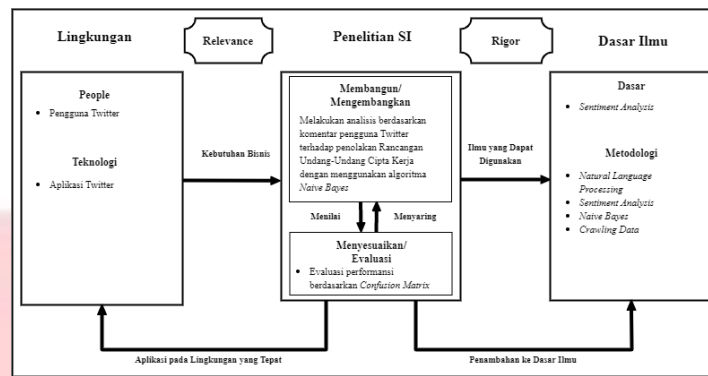
Accuracy adalah keberhasilan suatu algoritma dengan seberapa akurat sistem dapat mengklasifikasikan secara benar [15]. Berikut adalah rumus untuk menghitung *accuracy*:

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN}$$

3. Metodologi Penelitian

3.1 Model Konseptual

Model Konseptual adalah model yang digunakan untuk merelasikan kebutuhan dari permasalahan, memahami serta mengevaluasi suatu penelitian informasi. Adapun konsep model pada penelitian ini dapat dilihat pada Gambar 2 dibawah ini:

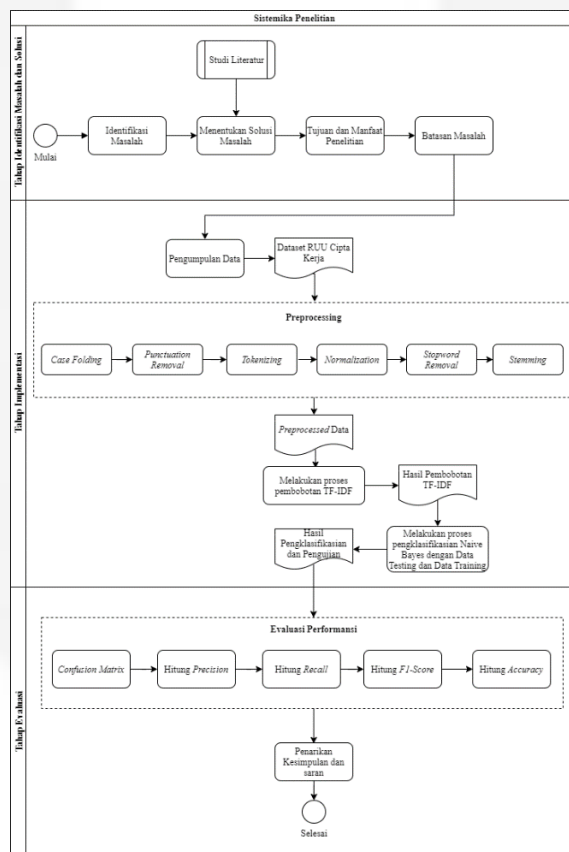


Gambar 2. Model Konseptual

Berdasarkan pada Gambar 2, pada lingkungan aspek *people* yaitu pengguna sosial media Twitter. Pengguna Twitter yang dimaksud adalah yang memberikan komentar terhadap adanya Rancangan Undang-Undang (RUU) Cipta Kerja dan pada aspek teknologi yaitu aplikasi Twitter itu sendiri. Penelitian ini memanfaatkan komentar masyarakat di teknologi aplikasi Twitter untuk dijadikan objek penelitian dan analisis sentimen. Penelitian ini mengkaji mengenai analisis sentimen media sosial Twitter terhadap kasus kampanye penolakan RUU cipta kerja menggunakan metode klasifikasi algoritma *Naive Bayes*. Untuk dasar ilmu yang digunakan adalah *sentiment analysis*. Kemudian, untuk metodologi yaitu *Natural Language Processing* (NLP) untuk mempermudah proses *preprocessing* terhadap *dataset* pada komentar masyarakat mengenai penolakan RUU cipta kerja, *text mining*, algoritma *Naive Bayes* dan *crawling data* untuk mengambil data dari Twitter. Kemudian, peneliti mengevaluasi algoritma klasifikasi *Naive Bayes* dengan melihat hasil *confusion matrix*, yang akan menghasilkan performansi evaluasi.

3.2 Sistematika Penelitian

Peneliti menggunakan sistematika penelitian dengan beberapa tahap, diantaranya yaitu tahap identifikasi masalah dan solusi, tahap implementasi dan tahap evaluasi. Diagram sistematika penelitian terjadi pada Gambar 3.



Gambar 3. Sistematika Penelitian

3.2.1 Tahap Identifikasi Masalah dan Solusi

Pada tahap pertama ini terlebih dahulu mengidentifikasi masalah yang terjadi mengenai reaksi masyarakat Indonesia terhadap adanya Rancangan Undang-Undang (RUU) Cipta Kerja di Indonesia. Kemudian, menentukan dari masalah yang diambil berdasarkan hasil studi literatur terkait identifikasi masalah. Selanjutnya, dari solusi masalah tersebut, peneliti menentukan tujuan dan manfaat masalah untuk memperjelas hal yang akan dicapai dari penelitian dan manfaatnya bagi sekitar. Peneliti menentukan batasan masalah agar penelitian lebih rinci dan tidak meluas.

3.2.2 Tahap Identifikasi Masalah dan Solusi

Pada tahap implementasi, peneliti melakukan pengumpulan data yang berupa kumpulan *tweets* dari pengguna Twitter yang berkaitan dengan studi kasus pada penelitian ini yaitu mengenai reaksi masyarakat pada Rancangan Undang-Undang (RUU) Cipta Kerja dan beberapa *hashtags* lain yang berkaitan dengan hal tersebut. Selanjutnya setelah *tweets* berhasil dikumpulkan dilanjutkan dengan melakukan *preprocessing*, yang diawali dengan melakukan proses *case folding*. Hal ini dilakukan untuk penyeragaman bentuk huruf dengan mengubah semua huruf menjadi huruf kecil untuk mempermudah proses *preprocessing* selanjutnya. Setelah itu, pada tahapan *punctuation removal* ini akan dilakukan penghapusan tanda baca yang terdapat pada kalimat komentar di *tweets*. Setelah melakukan tahap *punctuation removal*, peneliti melakukan tahapan *tokenizing* yang akan dilakukan proses pemotongan *string* input berdasarkan tiap kata yang menyusunnya. Kemudian, pada tahap *stopword removal* akan dilakukan penghapusan kata-kata yang tidak memiliki banyak kontribusi pada kalimat dan pada tahap selanjutnya peneliti melakukan *stemming*, agar komentar tidak memiliki kata imbuhan. Pada proses *preprocessing* ini akan menghasilkan data yang lebih terstruktur. Setelah dilakukan *preprocessing* selanjutnya dilakukan perhitungan pembobotan TF-IDF untuk mengetahui kata yang sering muncul dari *dataset*. Kemudian melakukan pemisahan *data training* dan *data testing* yang selanjutnya akan diklasifikasikan menggunakan algoritma *Naïve Bayes*. Tahap implementasi diakhiri dengan hasil klasifikasi analisis sentimen.

3.2.3 Tahap Evaluasi

Tahap evaluasi ini diawali dengan performansi evaluasi. Langkah pertama yaitu melakukan pemetaan hasil prediksi ke tabel *confusion matrix*. Dengan adanya *confusion matrix*, menghitung *precision*, *recall*, *f1-score* dan *accuracy* akan jauh lebih mudah. Tahapan selanjutnya adalah menghitung *precision* yaitu perbandingan jumlah prediksi benar dengan total prediksi. Setelah itu *recall* akan dihitung, yaitu perbandingan prediksi benar dengan total prediksi benar. Dengan adanya nilai *recall* dan *precision*, nilai *f1-score* akan bisa untuk dihitung. Tahapan performansi evaluasi diakhiri dengan menghitung akurasi dari model yang telah dibuat. Tahapan ini berakhir dengan menganalisis hasil dari performansi evaluasi, dan melakukan penarikan kesimpulan dan saran.

4. Pengujian dan Implementasi

4.1 Dataset

Dalam penelitian ini, *dataset* yang digunakan adalah komentar-komentar yang ada pada pengguna media sosial Twitter mengenai reaksi masyarakat pada Rancangan Undang-Undang (RUU) Cipta Kerja. Semua komentar pada *dataset* didapatkan dengan *crawling* data Twitter menggunakan *python* melalui aplikasi *jupyter notebook*. Komentar-komentar yang tidak berbahasa Indonesia dan komentar-komentar yang tidak relevan pada topik Rancangan Undang-Undang (RUU) Cipta Kerja akan dihapus, sehingga *dataset* yang terkumpul sebanyak 1667 data yang relevan. Kemudian, *dataset* dilakukan pelabelan secara manual berdasarkan sentimen dari *dataset* tersebut. Pelabelan data terdiri dari positif, negatif dan netral. Berikut ini pada Tabel 2 adalah contoh komentar hasil pelabelan.

Tabel 2. Contoh Pelabelan pada *Dataset*

Komentar	Label
Peraturan turunan UU Cipta Kerja akan memudahkan kemitraan usaha antara perusahaan besar dengan UMKM	Positif
Aturan turunan UU Cipta Kerja rupanya dinilai punya risiko menambah tekanan bagi UMKM di tengah pandemi Covid-19. Beberapa di antaranya ialah soal pesangon, investasi, dan perpajakan.	Negatif
Terlalu bnyk aturan.. Justru itu yg menciptakan/membuat bnyk permasalahan.. Maka timbul ide dn terobosan yg baru. (Uu cipta kerja.) gak perlu pake aturan napa mbak?	Negatif
Cukup banyak ketentuan mengenai lingkungan hidup yang diubah dalam UU Cipta Kerja, apa saja perubahannya?	Netral
Menciptakan lapangan pekerjaan adalah tujuan dari UU Cipta Kerja	Positif

4.2 Pembobotan TF-IDF

Pembobotan yang terdapat pada setiap kata yang ada di dalam dokumen harus dilakukan metode klasifikasi terhadap data yang berbentuk teks. Teknik pembobotan kata tersebut yaitu TF-IDF (*term frequency-inverse document frequency*.) Nilai TF-IDF ini dipengaruhi oleh kemunculan kata dalam sebuah dokumen dan jumlah kata atau frekuensi kata yang muncul secara keseluruhan. Berikut ini adalah contoh kalimat yang telah diproses pada tahapan *preprocessing* yang akan dihitung pembobotan kata “dukung uu cipta kerja maju ekonomi indonesia cipta lapangan kerja”, kalimat tersebut merupakan dokumen ke-210. Perhitungan nilai TF-IDF dapat dilihat pada Tabel 3.

Tabel 3. Perhitungan TF-IDF

Kata/Term	Urutan Tweet					IDF	TF-IDF				
	1	...	210	...	1667		1	...	210	...	1667
Dukung	0	...	0,1	...	0	3,101	0	...	0,3101	...	0
UU	0	...	0,1	...	0	0,498	0	...	0,0498	...	0
Cipta	0	...	0,2	...	0	0,526	0	...	0,1052	...	0
Kerja	0	...	0,2	...	0	0,497	0	...	0,0994	...	0
Maju	0	...	0,1	...	0	4,283	0	...	0,4283	...	0
Ekonomi	0	...	0,1	...	0	2,477	0	...	0,2477	...	0
Indonesia	0	...	0,1	...	0	2,100	0	...	0,2100	...	0
Lapang	0	...	0,1	...	0	3,506	0	...	0,3506	...	0

4.3 Hasil Klasifikasi Naïve Bayes

Berikut ini adalah *ratio* pada data *training* dan data *testing* yang dapat dilihat pada penjabaran Tabel 4.

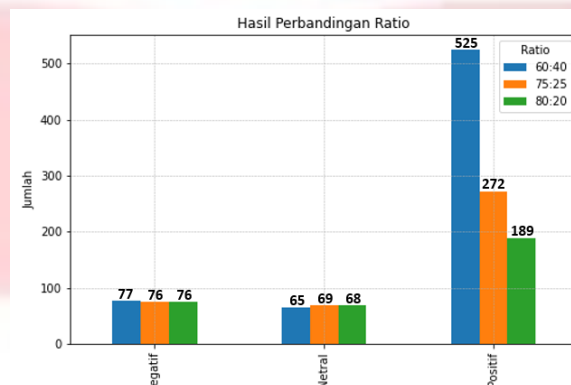
Tabel 4. *Ratio Testing*

Rasio	Hasil Akurasi	Data Training	Data Testing	Jumlah
60:40	79,16%	1000	667	1667
75:25	92,33%	1250	417	
80:20	90,39%	1334	333	

Dengan berbagai perbandingan rasio *testing*, model klasifikasi *Naïve Bayes* yang memiliki *performance* dengan akurasi tertinggi adalah perbandingan rasio 75:25 dengan nilai akurasi yaitu 92,33%. Oleh karena itu, peneliti akan menggunakan perbandingan rasio 75:25, dimana 75% data akan dijadikan data *training* dan 25% data akan dijadikan data *testing*. Berdasarkan perbandingan rasio tersebut dapat ditentukan bahwa untuk data *training* berjumlah 1250 data dan data *testing* berjumlah 417 data.

4.4 Analisis Sentimen & Evaluasi Performansi

Setelah membuat model klasifikasi menggunakan algoritma *naïve bayes*, kemudian melakukan analisis sentimen menggunakan data *testing* yaitu 417 data. Hasil prediksi oleh metode *Naïve Bayes* memberikan ketiga label positif, negatif dan netral sesuai dengan jumlah data *testing*. Hasil klasifikasi *Naïve Bayes* menunjukkan bahwa dengan model klasifikasi yang telah dibuat, menghasilkan klasifikasi sentimen seperti pada Gambar 5.



Gambar 5. *Bar Chart* Hasil Perbandingan Klasifikasi

Pada Gambar 5, dapat disimpulkan bahwa media sosial Twitter mengenai komentar terhadap reaksi masyarakat pada RUU Cipta Kerja adalah positif. Dengan ini, pengguna Twitter sebagian besar mendukung adanya mengenai adanya RUU Cipta Kerja di Indonesia. Komentar negatif disini tampak lebih banyak dari pada komentar netral. Komentar negatif diartikan sebagai komentar yang tidak mendukung atau menolak adanya RUU Cipta Kerja dan komentar netral diartikan bahwa sebagian masyarakat tidak mendukung maupun tidak menolak adanya RUU Cipta Kerja di Indonesia.

Kemudian, pada tahap evaluasi ini, setelah melakukan prediksi dengan data testing peneliti akan melakukan tahap evaluasi ini dengan *confusion matrix*. Berikut ini adalah hasil *confusion matrix* dari klasifikasi *Naïve Bayes*.

Tabel 5. Hasil Pengujian *Confusion Matrix*

<i>Predicted</i> \ <i>Actual</i>	Positif	Negatif	Netral	Total
Positif	225	39	8	272
Negatif	0	74	2	76
Netral	2	0	67	69
Total	227	113	77	417

Berdasarkan Tabel V-8, model klasifikasi *Naïve Bayes* pada data aktual sentimen positif didapatkan prediksi yang tepat sesuai dengan kategorinya dari data uji tersebut adalah sebesar 349 data. Artinya terdapat 349 data yang diprediksi oleh mesin atau model dengan tepat dan tidak terjadi *miss classification*. Tetapi dari total 378 data pada sentimen positif terdapat 24 data sentimen positif yang terprediksi pada sentimen negatif dan terprediksi 5 data pada sentimen netral. Pada kategori sentimen negatif terdapat 23 data yang diprediksi sesuai dengan data aktual, tetapi terdapat 1 data sentimen negatif yang terprediksi kedalam sentimen positif dan 0 data sentimen negatif yang terprediksi kedalam sentimen netral. Kemudian, pada kategori sentimen netral terdapat 13 data yang diprediksi sesuai dengan data aktual, tetapi terdapat 2 data sentimen netral yang terprediksi kedalam sentimen positif dan 0 data sentimen netral yang terprediksi kedalam sentimen negatif. Ketiga label positif, negatif dan netral dari hasil prediksi sebelumnya telah dipetakan pada *confusion matrix*. Selanjutnya, peneliti telah melakukan evaluasi performansi dari *confusion matrix* yang dibuat. Tabel 6 adalah hasil evaluasi performansi dengan total 417 jumlah data *testing*.

Tabel 6. Hasil Evaluasi Performansi

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Positif Label	82%	99%	90%
Negatif Label	97%	65%	78%
Netral Label	97%	87%	92%
Rata-Rata	92%	84%	86%
<i>Accuracy</i>	88%		

Pada Tabel 6, dapat dilihat bahwa nilai evaluasi performansinya dimana nilai *precision* yang didapat sebesar 92%, nilai *recall* sebesar 84% dan *F-1 score* sebesar 86%. Sehingga hasil evaluasi performansi dengan menggunakan rasio 75:25 mendapatkan akurasi yang cukup besar yaitu 88%.

4.5 Word Cloud

Pada penelitian ini dilakukan visualisasi data berupa *Word Cloud* yang berfungsi untuk memunculkan kata-kata yang sering digunakan oleh pengguna Twitter mengenai reaksi masyarakat pada RUU Cipta Kerja sehingga memudahkan pembaca dalam mengartikan atau memahami sebuah informasi. Hasil *Word Cloud* pada penelitian ini terdapat pada Gambar 6 berikut.



Gambar 6. Word Cloud

Dari hasil *word cloud* pada Gambar V.6 dapat dilihat kata “*omnibuslaw*”, “*cipta*”, “*kerja*” dan “*uu*” merupakan kata yang paling mendominasi jika dibandingkan dengan kata-kata lain. Hal ini dikarenakan dalam komentar Twitter yang membahas RUU Cipta Kerja, banyak yang menggunakan kata-kata tersebut. Selain itu, frekuensi kata-kata yang sering muncul juga ada *usaha*, *ekonomi*, *bangun*, *mudah* dan *dukung* yang menunjukkan sebagian masyarakat indonesia mendukung terhadap poin-poin yang ada pada RUU Cipta Kerja. Kemudian, frekuensi kata-kata yang juga sering muncul adalah kata *buruh*, *tolak*, *turun* yang menandakan bahwa kaum buruh/pekerja yang merasa keberatan terhadap poin-poin dalam RUU Cipta Kerja dan melakukan unjuk rasa dalam rangka penolakan terhadap adanya RUU Cipta kerja.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Analisis sentimen yang dilakukan menggunakan data komentar dari pengguna Twitter terhadap reaksi masyarakat pada RUU Cipta kerja dengan total 1667 komentar dan dilakukan *preprocessing* data yang terdiri dari tahap *case folding*, *punctuation removal*, *tokenization*, *stopword*, *normalization* dan *stemming*. Kemudian, melakukan klasifikasi *Naïve Bayes* dengan membagi data *training* dan data *testing* untuk mendapat prediksi melalui algoritma *Naïve Bayes*. Setelah itu dilakukan proses perhitungan evaluasi menggunakan metode *confusion matrix* untuk mendapatkan nilai *precision*, *recall*, *f1-score* dan *accuracy*.
2. Hasil performansi evaluasi analisis sentimen terhadap reaksi masyarakat pada RUU Cipta Kerja menggunakan klasifikasi algoritma *Naïve Bayes* ini memiliki akurasi yang paling baik dengan mengambil rasio *testing* 75:25 sebagai acuan terbaik dalam melakukan prediksi data *training* sebanyak 1250 data dan data *testing* sebanyak 417 data. Hasil evaluasi performansi ini terdapat nilai *recall*, *precision* dan *F1-Score* dengan memiliki nilai rata-rata yang baik yaitu untuk *recall* memiliki rata-rata nilai sebesar 84%, *Precision* sebesar 92% dan *F1-Score* sebesar 86%. Menurut peneliti, dengan nilai rata-rata tersebut bernilai diatas 80%, peneliti menarik kesimpulan bahwa performa klasifikasi pada algoritma *Naïve Bayes* sudah cukup baik. Hal ini juga dibuktikan dengan nilai akurasi yang cukup tinggi yaitu 88%.
3. Dampak reaksi masyarakat pada RUU Cipta Kerja terhadap kecenderungan komentar pengguna Twitter adalah positif. Komentar positif disini adalah pengguna Twitter lebih banyak mendukung dengan adanya RUU Cipta Kerja di Indonesia. Hal ini dibuktikan dengan banyaknya frekuensi kata seperti *usaha*, *ekonomi*, *bangun*, *mudah*, *dukung* yang menunjukkan sebagian masyarakat indonesia mendukung terhadap poin-poin yang ada pada RUU Cipta Kerja, karena pengguna media sosial Twitter menganggap bahwa RUU Cipta Kerja ini dapat membangun NKRI dalam memudahkan percepatan seperti di sektor pertanian dan mempercepat pertumbuhan ekonomi di indonesia.

REFERENSI

- [1] Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), 4704–4713.
- [2] Gunawan, D., Riana, D., Ardiansyah, D., & Akbar, F. (2020). Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. *Jurnal Teknik Komputer*, 6(1), 121–129.
- [3] Mayasari, I. (2020). KEBIJAKAN REFORMASI REGULASI MELALUI IMPLEMENTASI OMNIBUS LAW DI INDONESIA. *Jurnal Rechts Vinding: Media Pembinaan Hukum Nasional*, 9(1), 1.
- [4] Kurniawan, F. (2020). PROBLEMATIKA PEMBENTUKAN RUU CIPTA KERJA DENGAN KONSEP OMNIBUS LAW PADA KLASSTER KETENAGAKERJAAN PASAL 89 ANGKA 45 TENTANG PEMBERIAN PESANGON KEPADA PEKERJA YANG DI PHK. *Jurnal Panorama Hukum*, 5(1)

- [5] Hadna, M.S., Santosa, P.I., and Winarno, W.W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. Seminar Nasional Teknologi Informasi Dan Komunikasi 2016, 57-64.
- [6] Agarwal, A. et al., 2014. Sentiment Analysis of Twitter Data. Department of Computer Science Columbia University.
- [7] Idris, M. (2019). IMPLEMENTASI DATA MINING DENGAN ALGORITMA NAÏVE BAYES UNTUK MEMPREDIKSI ANGKA KELAHIRAN, *Jurnal Pelita Informatika*, 7(3), 421-428.
- [8] Utomo, D. P., & Purba, B. (2019). Penerapan Data mining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 846.
- [9] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50.
- [10] Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 306-312.
- [11] Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 306-312.
- [12] Rustiana, D., & Rahayu, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 8(1), 113-120.
- [13] Simorangkir, H., & Lhaksana, K. M. (2018). Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier. *e-Proceeding of Engineering* , 8131-8140.
- [14] Sasongko, T. B. (2016). Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA). *Jurnal Teknik Informatika dan Sistem Informasi*, 244-253.
- [15] Agasty, I. M. (2018). PENGARUH STEMMER BAHASA INDONESIA TERHADAP PERORMA ANALISIS SENTIMEN TERJEMAHAN ULASAN FILM. *Jurnal TEKNOKOMPAK*, 18-23.