



Analisis Sentimen Opini Publik Terhadap Undang-Undang Cipta Kerja pada Twitter Menggunakan Metode *Naive Bayes Classifier*

(Sentiment Analysis of Public Opinion Against the Job Creation Law from Twitter

Using The Naïve Bayes Classifier Method)

Yanuar Nurdiansyah^{1*}, Fatchur Rahman², Priza Pandunata³, Ardian Infantono⁴

^{1,2,3} Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Jember

E-mail: ¹yanuar_pssi@unej.ac.id, ²172410102033@mail.unej.ac.id

⁴ Program Studi Teknologi Aeronautika Pertahanan, Akademi Angkatan Udara, Yogyakarta

E-mail: ⁴ardian.infantono@aau.ac.id

Abstract— *Sentiment analysis or Opinion Mining is a way of solving a problem based on public opinion that is widely circulated on social media which is expressed in text form. Sentiment analysis is very helpful for the government / an agency in knowing public opinion about a policy that has just been issued without using conventional survey methods. The sentiment analysis carried out focuses on trending tweet topics on Twitter with trending topics on October 5 to 10 are #Omnibuslaw, #tolakruuciptakerja, #UUCiptaKerja, and #tolakomnibuslaw, and the trending topic on November 21 and 22 is "obl makmurkan buruh". The sentiment analysis process is carried out after the data is obtained at the data crawling stage, followed by word cleaning in the preprocessing process, and word weighting with the TF-IDF algorithm. Sentiment analysis using the naive bayes classifier method aims to obtain a classification of public opinion on the job creation law on twitter. There are two classes in this study, there are positive and negative classes. The 2000 dataset consisting of 1400 tweets that have negative sentiments & 600 positive tweets used will be divided between training data and testing data with a ratio of 60%: 40%, 70%:30%, 80%:20%, and 90 %:10%. From the evaluation results on sentiment analysis regarding public opinion on the copyright law on Twitter, the highest accuracy value is 94% with training data used at 90%, testing data at 10%. In its implementation, the results of the sentiment test show that negative sentiment results are higher than positive sentiment.*

Keywords— *Sentiment analysis, job creation law, TF-IDF algorithm, naive bayes classifier method*

Abstrak— *Analisis sentiment atau Opinion Mining merupakan cara memecahkan suatu permasalahan berdasarkan opini masyarakat yang beredar luas di media sosial yang diekspresikan dalam bentuk teks. analisis sentimen sangat membantu pemerintahan/ suatu instansi dalam mengetahui opini publik mengenai suatu kebijakan yang baru saja dikeluarkan tanpa menggunakan metode survey konvensional. Pada analisis sentimen yang dilakukan berfokus pada Trending topik tweet pada Twitter dengan trending topic pada tanggal 5 sampai 10 oktober yaitu #Omnibuslaw, #tolakruuciptakerja, #UUCiptaKerja, dan #tolakomnibuslaw, dan trending topic pada tanggal 21 dan 22 november yaitu "obl makmurkan buruh". Proses Analisis sentimen dilakukan setelah data didapatkan pada tahapan crawling data, dilanjutkan dengan pembersihan kata pada proses preprocessing, dan pembobotan kata dengan algoritma TF-IDF. Analisis sentimen menggunakan metode naive bayes classifier bertujuan agar mendapatkan klasifikasi mengenai opini publik terhadap undang-undang cipta kerja pada twitter. Terdapat dua kelas pada penelitian ini yaitu kelas positif, dan negatif. Dari 2000 dataset yang terdiri dari 1400 tweet yang bersentimen negatif & 600 tweet yang bersifat positif dipakai akan dibagi*

*Penulis Korespondensi (Yanuar Nurdiansyah)

E-mail: yanuar_pssi@unej.ac.id

antara data training dan data testing dengan perbandingan sebesar 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Dari hasil evaluasi pada Analisis sentimen mengenai opini publik terhadap undang-undang cipta kerja pada twitter didapatkan nilai akurasi tertinggi sebesar 94% dengan data training yang dipakai sebesar 90%, data testing sebesar 10%. Pada implementasinya, hasil dari uji sentimen menunjukkan hasil sentimen negatif yang lebih tinggi dibandingkan sentimen positif.

Kata Kunci— Analisis Sentimen, Undang-undang Cipta Kerja, Algoritma TF-IDF, Naive Bayes Classifier

I. PENDAHULUAN

Pada setiap satu periode masa jabatan presiden tentunya pasti ada pembuatan Undang-undang baru guna mengkaji ulang Undang-undang yang ada pada masa jabatan presiden sebelumnya dan salah satu Undang-undang yang telah diresmikan pada masa jabatan presiden Joko Widodo adalah Undang-undang cipta kerja [1]. Pembuatan Undang-undang cipta kerja ini merupakan bagian dari program kerja selama satu periode yang sebelumnya pada Sidang Paripurna MPR RI dalam rangka kegiatan Pelantikan Presiden beserta Wakil Presiden telah menyatakan bahwa akan terdapat penyederhanaan regulasi yang akan beliau kerjakan. Hal ini sudah terealisasi pada pasal 1 ayat (1) Undang-Undang Republik Indonesia tentang cipta kerja tahun 2020, dijelaskan bahwa cipta kerja merupakan sebuah upaya dalam membuka lapangan kerja baru melalui usaha kemudahan, perlindungan, pemberdayaan koperasi & UMKM, kemudahan dalam membuka wirausaha, dan meningkatkan ekosistem investasi serta percepatan proyek strategis nasional di Indonesia. Pengesahan RUU cipta kerja tentunya menimbulkan dampak positif dan negatif. Pemberitaan ini tentunya menjadi perdebatan pro dan kontra dikalangan masyarakat terkait pengesahan RUU cipta kerja, hal ini dikarenakan pengesahan undang-undang cipta kerja dinilai terlalu cepat dan masih ada beberapa pasal yang menurut masyarakat kurang tepat sasaran dan tidak sesuai dengan kejadian di lapangan. Untuk mempresentasikan aspirasi masyarakat mengenai fenomena tersebut banyak media yang dapat digunakan, keberadaan media sosial tentunya menjadi pusat penyebaran informasi dan opini pada masa pandemi COVID-19 [2].

Media sosial pada awalnya bertujuan mempermudah individu satu dengan individu lainnya dalam berkomunikasi yang sudah terpisah namun masih merasa membutuhkan komunikasi satu sama lain. Namun pada perkembangannya, media sosial justru digunakan publik sebagai ajang penyebaran informasi & opini [3]. Media sosial tentunya memberikan kebebasan bagi individu dalam mengutarakan argumen yang tidak bisa mereka ungkapkan secara langsung. Di sisi lain, penggunaan media sosial seringkali meresahkan di lingkungan sekitar, dimana kebebasan beropini disalahgunakan secara semena-mena tanpa mempertimbangkan Batasan aturan & norma yang berlaku dimasyarakat [4]. Salah satu media sosial yang populer pada masa pandemi COVID-19 adalah *Twitter*.

Twitter adalah sebuah mikroblog dan juga media sosial yang memudahkan penggunaanya dalam berkiriman dan juga membaca pesan (*tweet*) yang jumlahnya tidak lebih dari 280 karakter [5]. *Twitter* menjadi sosial media yang masih eksis meskipun disaingi oleh media social lainnya dikarenakan kecepatan arus informasi yang didapat oleh pengguna dan kemudahan dalam berinteraksi dengan sesama pengguna, hal ini dibuktikan dengan adanya fitur yang tidak dimiliki media sosial yang lain yaitu fitur *trending topic*. *Trending topic* merupakan daftar topik/ tren yang disesuaikan untuk pengguna berdasarkan suatu yang sedang hangat dibicarakan disuatu wilayah [6]. *Twitter* tentunya digunakan oleh pemerintah untuk berkomunikasi dengan masyarakat perihal adanya kebijakan baru di Indonesia. Hal tersebut dikarenakan kecepatan arus informasi pada fitur *trending topic* yang dijadikan pemerintah sebagai saluran komunikasi politik dengan jumlah pengguna *twitter* Indonesia mencapai 19,5 juta [6].

Analisis sentimen atau *Opinion Mining* merupakan cara memecahkan suatu permasalahan berdasarkan opini masyarakat yang beredar luas di media sosial yang diekspresikan dalam

bentuk teks [7]. Analisis sentiment dapat digunakan sebagai alat bantu melihat sentimen masyarakat mengenai suatu kejadian tertentu, sehingga dapat segera dipakai sebagai bahan evaluasi kejadian dimasa mendatang [8]. Dewasa ini, analisis sentimen sangat membantu pemerintahan/ suatu instansi dalam mengetahui opini publik mengenai suatu kebijakan yang baru saja dikeluarkan tanpa menggunakan metode survei konvensional. Analisis sentimen akan mengklasifikasikan polaritas dari suatu teks yang terdapat pada tweet untuk mengetahui opini yang ada dalam tweet tersebut bersifat negatif, atau positif [9].

II. LANDASAN TEORI

A. Algoritma Term Frequency-Inverse Document Frequency (TF-IDF)

Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) merupakan algoritma penggabungan dua konsep dalam menghitung sebuah bobot kata, yaitu inverse frekuensi suatu dokumen yang terdapat pada kata tersebut & frekuensi kemunculan sebuah kata pada dokumen tersebut. Hasil frekuensi kemunculan kata yang terdapat pada setiap dokumen akan dijadikan sebagai indikator seberapa pentingnya kata itu yang terdapat pada dokumen tersebut. Hasil frekuensi dokumen yang terdapat kata tersebut akan menentukan tolak ukur seberapa umum kata tersebut[10].

Term Frequency adalah frekuensi jumlah kemunculan sebuah kata pada suatu dokumen yang bersifat teks. Nilai TF akan dideklarasikan sebagai total dari kemunculan term pada sebuah dokumen. *Document Frequency* adalah jumlah kemunculan kata yang ada pada banyak dokumen. Kata tersebut bisa informatif maupun tidak informatif, seperti kata ke, tetapi, kalau, jika, kenapa, adalah, dll.

Invers Document Frequency (IDF) merupakan hasil perhitungan dari frekuensi kemunculan suatu term pada seluruh dokumen yang bersifat teks. Nilai term yang sering muncul pada seluruh dokumen tersebut memiliki nilai *Invers Document Frequency (IDF)* lebih kecil dibandingkan dengan term yang jarang muncul [11]. *Invers Document Frequency (IDF)* dirumuskan dengan persamaan (1) sebagai berikut :

$$Idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right) \quad (1)$$

Setelah mendapatkan nilai TF, dan IDF, maka selanjtnya akan mulai menentukan nilai TF-IDF. nilai TF-IDF merupakan hasil kombinasi dari perkalian dari nilai *Term Frequency (TF)* dan nilai *Invers Document Frequency (IDF)* dalam menghitung nilai TF-IDF. TF-IDF dirumuskan dengan persamaan (2) sebagai berikut :

$$W_{t,d} = W_{tf,t,d} \times Idf_t \quad (2)$$

B. Metode Naive Bayes Classifier (NBC)

Naive Bayes Classifier merupakan sebuah algoritma klasifikasi yang bersumber pada pada teorema bayes dimana pada proses klasifikasi yang dilakukan dengan melakukan perhitungan nilai probabilitas dari suatu data yang berbentuk teks & dapat mengolah data dengan jumlah yang besar dengan nilai akurasi yang tinggi [12]. Persamaan dari teorema Bayes dirumuskan menggunakan persamaan (3) sebagai berikut :

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)} \quad (3)$$

Cara kerja dari algoritma *Naive Bayes Classifier* yaitu dengan melakukan perhitungan nilai *prior probability*, *nilai conditional probability*, dan juga proses perhitungan hasil prediksi. *prior probability* merupakan nilai probabilitas dari pembagian jumlah dokumen pada suatu kelas terhadap jumlah keseluruhan dokumen. Rumus perhitungan *prior probability* dirumuskan menggunakan persamaan (4) sebagai berikut:

$$P(V_j) = \frac{D_j}{D} \quad (4)$$

Bobot term pada semua kelas prediksi dapat dihitung menggunakan *conditional probability* yang merupakan nilai bobot kata terhadap semua kelas prediksi. *Rumus conditional probability* dirumuskan menggunakan persamaan (5) sebagai berikut:

$$P(X_1|V_j) = \frac{W_{ij}+1}{N+N_j} \quad (5)$$

Pada tahap selanjutnya dilakukan perhitungan hasil prediksi yang merupakan nilai probabilitas suatu dokumen pada kelas tertentu. Rumus dari perhitungan hasil prediksi probabilitas tertinggi dirumuskan menggunakan persamaan (6) sebagai berikut :

$$V_{\text{map}} = V_j e^{\text{Var}gmax} \frac{P(V_j)P(X_1, X_2, X_3, \dots, X_n)}{P(X_1, X_2, X_3, \dots, X_n)} \quad (6)$$

Jika hasil dari $P(X_1, \dots, X_n)$ bernilai konstan pada seluruh kelas V_j , maka persamaan nya akan diganti menjadi persamaan (7) sebagai berikut:

$$V_{\text{map}} = V_j e^{\text{Var}gmax} P(X_1|V_j) P(V_j) \quad (7)$$

Sehingga persamaannya disederhanakan menjadi persamaan (8) sebagaimana berikut:

$$V_{\text{map}} = \prod_{i=1}^n P(X_i|V_j) P(V_j) \quad (8)$$

:

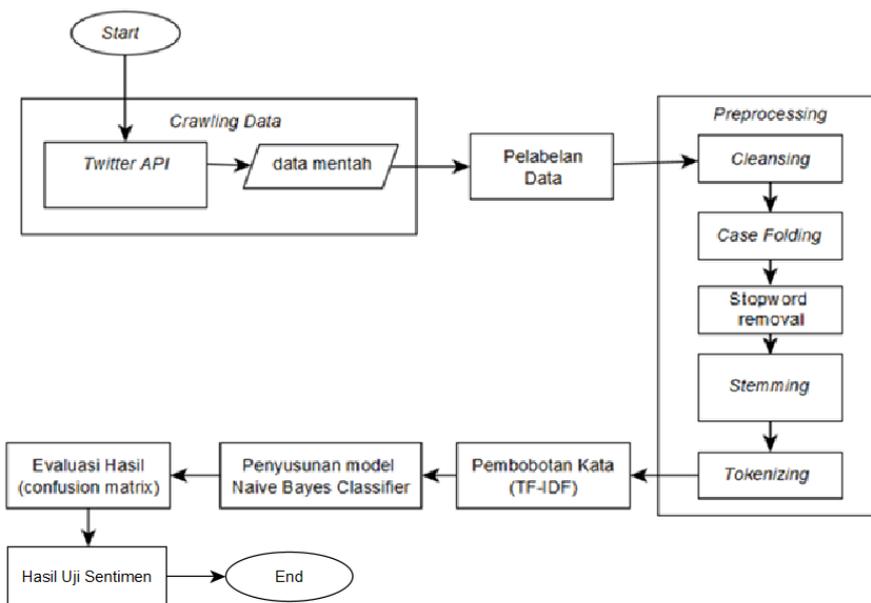
III. MODEL YANG DIUSULKAN

A. Jenis Penelitian

Jenis dari penelitian ini adalah penelitian terapan. Penelitian terapan adalah penelitian yang dipergunakan sebagai pemecah masalah yang praktis, baik individu maupun kelompok. Tujuan dari penelitian terapan yaitu untuk memberikan solusi atas suatu permasalahan tertentu dihadapi oleh masyarakat, atau suatu organisasi / instansi pemerintahan [13]. Penelitian ini akan berfokus pada konsekuensi yang akan muncul dari sebuah kebijakan mengenai UU cipta kerja.

B. Object Penelitian

Objek dari penelitian yang dipakai adalah *Tweet* dari media sosial *Twitter* yang membahas tentang UU cipta kerja. Data yang didapatkan untuk melakukan penelitian diperoleh langsung dari Trending topic Twitter yaitu #Omnibuslaw, #tolakruuciptakerja, #UUCiptaKerja, #tolakomnibuslaw, dan "obl makmurkan buruh" menggunakan Teknik *crawling data* dengan *Twitter API key* dari penulis. Hasil *crawling data* akan didapatkan 2000 data *tweet*. Tahapan yang dilakukan pada penelitian ini digambarkan melalui gambar 1 sebagai berikut:



Gambar 1. Diagram alir PC dengan GPS

IV. IMPLEMENTASI MODEL DAN PEMBAHASAN

Sebelum mendapatkan dataset dibutuhkan kode Twitter API Key guna mendapatkan data tweet pada twitter yang digunakan. Pada penelitian ini pengambilan Twitter API Key dilakukan dengan cara membuat akun sebagai developer twitter pada situs developer twitter agar bisa mendapatkan kode Twitter API Key untuk melakukan proses crawling data. Pada saat menggunakan Twitter API Key, maka dibutuhkan beberapa kode untuk akses Twitter API Key seperti consumer key, consumer secret, access token secret, dan access token.

Pengambilan dataset pada twitter akan dipakai untuk data training dan data testing. Data training dipakai guna bahan pembelajaran pada tahap klasifikasi sehingga hasil klasifikasi dapat menguji data testing. Dataset diperoleh dari crawling data dan akan diberikan label secara manual. Pelabelan dataset dilakukan secara manual oleh ahli Bahasa yaitu alumnus program studi Pendidikan Bahasa dan Sastra Bahasa Indonesia, Fakultas Keguruan dan Ilmu Pendidikan Universitas Jember. Penelitian ini akan membagi dataset dibagi menjadi data training dan data testing dengan perbandingan sebesar 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Kelas klasifikasi penelitian ini yaitu: positif, dan negatif. Pada kelas yang bersifat positif data berjumlah 600 tweet. Sedangkan kelas yang bersifat negatif berjumlah 1400 tweet. Sehingga jumlah keseluruhan dataset pada penelitian ini sebanyak 2000 tweet. Contoh dataset yang dipakai bisa dilihat pada Tabel I sebagai berikut:

TABEL I
CONTOH DATASET

No	Data Training	Sentimen
1	Uu cipta kerja mendorong umkm di tengah pademi covid19 #uuciptakerja #UUCiptaKerjaHarapanRakyat #UUCiptaker	Positif
2	biarkan polisi sibuk sendiri kita tetap #TolakOmnibusLaw	Negatif

Pada tahap preprocessing dataset diawali dengan tahap cleansing yaitu menghapus karakter kosong, angka, hastag, symbol, url, dan emoticon. Hasil cleansing dapat dilihat pada Tabel II. sebagai berikut:

TABEL II
CONTOH HASIL CLEANSING

Tweet	Hasil Cleansing
Uu cipta kerja mendorong umkm di tengah pademi covid19 #uuciptakerja #UUCiptaKerjaHarapanRakyat #UUCiptaker	Uu cipta kerja mendorong umkm di tengah pademi covid uuciptakerja UUCiptaKerjaHarapanRakyat UUCiptaker
biarkan polisi sibuk sendiri kita tetap #TolakOmnibusLaw	biarkan polisi sibuk sendiri kita tetap TolakOmnibusLaw

Setelah tahap cleansing selesai, maka dilanjutkan dengan tahap *casefolding*. Pada tahap *casefolding*, semua kalimat pada dataset akan diubah menjadi *lowercase*. Hasil *casefolding* dapat dilihat pada tabel III. sebagai berikut:

TABEL III
CONTOH HASIL CASEFOLDING

No	Cleansing	Casefolding
1	Uu cipta kerja mendorong umkm di tengah pademi covid uuciptakerja UUCiptaKerjaHarapanRakyat UUCiptaker	uu cipta kerja mendorong umkm di tengah pademi covid uuciptakerja uuciptakerjahaanrakyat uuciptaker
2	biarkan polisi sibuk sendiri kita tetap TolakOmnibusLaw	biarkan polisi sibuk sendiri kita tetap tolakomnibuslaw

Setelah menyelesaikan tahap *casefolding*, maka dilanjutkan ke tahap *stopword remover*. Proses *stopword remover* akan menghapus kata yang tidak penting sehingga dapat mempercepat pemrosesan dataset karena hanya mengambil kata yang memiliki makna penting pada dataset. hasil *stopword remover* dapat dilihat pada Tabel IV. sebagai berikut:

TABEL IV
CONTOH HASIL STOPWORD REMOVER

No	Casefolding	Stopword remover
1	uu cipta kerja mendorong umkm di tengah pademi covid uuciptakerja uuciptakerjahaanrakyat uuciptaker	uu cipta kerja mendorong umkm tengah pademi covid uuciptakerja uuciptakerjahaanrakyat uuciptaker
2	biarkan polisi sibuk sendiri kita tetap tolakomnibuslaw	biarkan polisi sibuk sendiri tetap tolakomnibuslaw

Setelah dataset sudah menyelesaikan proses *stopword remover*, maka selanjutnya akan melakukan proses *stemming* yaitu mengganti kata yang memiliki imbuhan menjadi kata yang tidak memiliki imbuhan / kata dasar. Dengan mengimplementasikan proses *stemming*, maka kata yang awalnya memiliki imbuhan awalan maupun akhiran akan dikembalikan ke kata dasar. Contoh hasil dari proses *stemming* dapat dilihat pada Tabel V. sebagai berikut:

TABEL V
CONTOH HASIL STEAMING

No	Stopword remover	Stemming
1	uu cipta kerja mendorong umkm tengah pademi covid uuciptakerja uuciptakerjahaanrakyat uuciptaker	uu cipta kerja dorong umkm tengah pademi covid uuciptakerja uuciptakerjahaanrakyat uuciptaker
2	biarkan polisi sibuk sendiri tetap tolakomnibuslaw	biar polisi sibuk sendiri tetap tolakomnibuslaw

Proses *stemming* selesai maka dilanjutkan dengan proses *tokenizing* yaitu proses memisahkan kalimat menjadi kata atau token. Kalimat dipisah berdasarkan spasi yang terdapat pada kalimat. Contoh hasil proses *tokenizing* dapat dilihat pada Tabel VI. sebagai berikut:

TABEL VI
CONTOH HASIL TOKENIZING

No	Stemming	Tokenizing
1	hoax fakta uu omnibus law tolak demo padahal banyak baik uu omnibuslaw jokowi indonesiamaju	['hoax', 'fakta', 'uu', 'omnibus', 'law', 'tolak', 'demo', 'padahal', 'banyak', 'baik', 'uu', 'omnibuslaw', 'jokowi', 'indonesiamaju']
2	uu cipta kerja dorong umkm tengah pademi covid uuciptakerja uuciptakerjahanrakyat uuciptaker	['uu', 'cipta', 'kerja', 'dorong', 'umkm', 'tengah', 'pademi', 'covid', 'uuciptakerja', 'uuciptakerjahanrakyat', 'uuciptaker']

Pembobotan kata mengimplementasikan algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)* dimulai pada tahap menentukan nilai *Term Frequency (TF)* dengan menentukan frekuensi jumlah kemunculan sebuah kata dari suatu dokumen yang bersifat teks dan mengimplementasikan kode program nilai *Term Frequency (TF)*. Contoh implementasi perhitungan *Term Frequency (TF)* yang sudah melewati tahap *preprocessing* dapat dilihat pada Tabel VII. sebagai berikut:

TABEL VII
CONTOH HASIL PERHITUNGAN TERM FREQUENCY (TF)

No	Dataset bersih	Term Frequency (TF)	
1	['uu', 'cipta', 'kerja', 'dorong', 'umkm', 'tengah', 'pademi', 'covid', 'uuciptakerja', 'uuciptakerjahanrakyat', 'uuciptaker']	kata	Nilai TF
		Uu	1/11
		cipta	1/11
		kerja	1/11
		dorong	1/11
		umkm	1/11
		tengah	1/11
		pademi	1/11
		covid	1/11
		uuciptakerja	1/11
		uuciptakerja harapanrakyat	1/11
uuciptaker	1/11		
2	['biar', 'polisi', 'sibuk', 'sendiri', 'tetap', 'tolakomnibuslaw']	kata	Nilai TF
		biar	1/7
		polisi	1/7
		sibuk	1/7
		sendiri	1/7
		tetap	1/7
		tolakomnibusl	1/7
		aw	

Setelah menyelesaikan tahap perhitungan *Term Frequency (TF)*, maka dilanjutkan dengan melakukan proses menentukan nilai *document frequency (DF)* dan menentukan nilai *invers document frequency (IDF)* dengan menghitung nilai log terhadap hasil *document frequency(DF)*. Contoh hasil nilai *document frequency(DF)* dan *invers document frequency (IDF)* dapat dilihat pada Tabel VIII. sebagai berikut:

TABEL VIII
CONTOH HASIL NILAI DOCUMENT FREQUENCY (DF) DAN INVERS DOCUMENT FREQUENCY (IDF)

No	Kata	Nilai DF	$Idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right)$
1	uu	430	0.667555677069
	cipta	529	0.577572217295
	kerja	620	0.508637437576
	dorong	13	2.18708649135
	umkm	40	1.69897000434
	tengah	28	1.85387195564
	pandemi	47	1.62893120399
	covid	44	1.65757688488
	uuciptakerja	260	0.886056213398
	uuciptakerjatarapan rakyat	8	2.39794000867
	uuciptaker	5	2.60205999133
2	biar	22	1.95860688055
	polisi	36	1.74472315193
	sibuk	13	2.18708649135
	sendiri	32	1.79588001734
	tetap	34	1.76955086147
	tolakomnibuslaw	322	0.793168521533

Pada proses selanjutnya yaitu perhitungan bobot TF-IDF yaitu mengkalikan hasil *Term Frequency (TF)* dengan *invers document frequency (IDF)*, sehingga hasil pembobotan TF-IDF dapat dilihat pada Tabel IX. sebagai berikut:

TABEL IX
CONTOH HASIL PEMBOBOTAN TF-IDF

No	Kata	$W_{t,d} = W_{tf_t} \times Idf_t$
1	uu	0.667555677069
	cipta	0.577572217295
	kerja	0.508637437576
	dorong	2.18708649135
	umkm	1.69897000434
	tengah	1.85387195564
	pandemi	1.62893120399
	covid	1.65757688488
	uuciptakerja	0.886056213398
	uuciptkerjatarapanrakyat	2.39794000867
	uuciptaker	2.60205999133

TABLE IX Lanjutan

No	Kata	$W_{t,d} = W_{tf,t} \times Idf_t$
2	biar	1.95860688055
	polisi	1.74472315193
	sibuk	2.18708649135
	sendiri	1.79588001734
	tetap	1.76955086147
	tolakomnibuslaw	0.793168521533

Penyusunan model dengan menerapkan metode *Naive Bayes Classifier* merupakan *training* data yang diawali dengan menentukan nilai *prior probability*. Contoh perhitungan nilai *prior probability* dapat dilihat pada Tabel X. sebagai berikut:

TABEL X
CONTOH HASIL NILAI PRIOR PROBABILITY

Sentimen	$P(v_j) = \frac{D_j}{D}$
positif	0.30666667
negatif	0.69333333

Tabel 10. merupakan contoh hasil nilai *prior probability* dengan membagi jumlah seluruh dokumen pada sebuah kelas dengan jumlah dataset yang terdapat pada data training. Hasil dari nilai perhitungan *prior probability* akan digunakan untuk proses menentukan hasil perhitungan suatu prediksi. Langkah selanjutnya yaitu menghitung nilai dari *conditional probability* yaitu menentukan nilai bobot kata terhadap semua kelas prediksi. Contoh perhitungan *conditional probability* dapat dilihat pada Tabel XI. sebagai berikut:

TABEL XI
CONTOH HASIL TRAINING DATA

Sentimen	Kata	$P(X_1 positif) = \frac{W_{(X_1 positif)}}{N + N_j}$	$\frac{W_{(X_1 negatif)}}{N + N_j} = \frac{W_{ij} + 1}{N + N_j}$
positif	uu	0.0003950617	0.00038879824
	cipta	0.0003661109	0.00036687725
	kerja	0.0003472122	0.00034641502
	dorong	0.0007863524	0.00078888279
	umkm	0.0006615122	0.00066806188
	tengah	0.0007044857	0.00070344391
	pandemi	0.0006470418	0.00064688267
	covid	0.0006542532	0.00065425329
	uuciptakerja	0.0004605753	0.00044430063
	uuciptakerjahar apanrakyat	0.0008404501	0.00084003461
	uuciptaker	0.0008911578	0.00089093742

TABLE XI Lanjutan

Sentimen	Kata	$P(X_1 positif)$	$\frac{W(x_1 negatif)}{N + N_j}$	$\frac{W_{ij} + 1}{N + N_j}$
negatif	biar	0.0007314232	0.00072925976	
	polisi	0.00067921879	0.00067355169	
	sibuk	0.00078849245	0.00078674067	
	sendiri	0.00069102323	0.0006876242	
	tetap	0.0006826598	0.0006826598	
	tolakomnibusla		0.0004419937	
	w	0.00044199371		

Setelah menyelesaikan dan mendapatkan nilai dari *conditional probability*, maka selanjutnya akan melakukan tahap *testing*. pada tahap ini *Naive Bayes Classifier* menentukan nilai probabilitas tertinggi dari kedua kelas target. Contoh hasil *testing* data dapat dilihat pada Tabel XII. sebagai berikut:

TABEL XII
CONTOH HASIL TESTING DATA

Sentimen Prediksi	Dokumen	$V_{MAP} = \prod_{i=1}^n P(X_i positif)P(V_j)$	$V_{MAP} = \prod_{i=1}^n P(X_i negatif)P(V_j)$	Kelas Target
Positif	Uu cipta kerja mendorong umkm di tengah pademi covid19 #uuciptakerja #UUCiptaKerjaHarapanRakyat #UUCiptaker	0.000588758	0.000445597	Positif
Negatif	biarkan polisi sibuk sendiri kita tetap #TolakOmnibusLaw	0.00066733	0.00088256	Negatif

Proses uji evaluasi menggunakan *confusion matrix* guna mengetahui tingkat akurasi proses klasifikasi yang sudah dikerjakan oleh sistem dengan acuan pada nilai akurasi. Pada penelitian ini, penulis melakukan empat kali percobaan dengan perbandingan antara *data training* dengan *data testing* sebesar 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. pada implementasinya, penulis menggunakan perhitungan akurasi, *precision*, *recall* dan *f-measure* dan hasilnya dapat dilihat pada Tabel XIII. sebagai berikut:

TABEL XIII
CONTOH HASIL UJI EVALUASI

Keterangan	Percobaan 1	Percobaan 2	Percobaan 3	Percobaan 4
Data Train	60%	70%	80%	90%
Data Testing	40%	30%	20%	10%
True Positif (TP)	202	150	103	53
True Negatif (TN)	541	401	268	135
False Positif (FP)	38	30	17	7
False Negatif (FN)	19	19	12	5
Akurasi	92.87	91.83	92.75	94
Presisi	0.84	0.89	0.9	0.91
Recall	0.91	0.83	0.86	0.88
F-Measure	87.36	85.89	85.89	87.95

Dari keempat percobaan yang dilakukan pada proses uji evaluasi menggunakan *confusion matrix* didapatkan hasil terbaik pada percobaan yang pertama dengan data training yang dipakai sebesar 90%, data testing sebesar 10% dan hasil akurasi sebesar 94%. Evaluasi uji sentimen dilakukan agar dapat mengetahui sentimen dari tweet yang telah diuji mengandung sentimen yang positif/negatif. Pada tahap uji sentimen, data tweet yang digunakan didapatkan dari hasil crawling data diluar tanggal pengambilan dataset. Proses uji sentimen dilakukan dengan melakukan empat kali percobaan dengan perbandingan antara data training dengan data testing sebesar 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Contoh hasil percobaan uji sentimen dapat dilihat pada tabel XIV sebagai berikut:

TABEL XIV
CONTOH HASIL UJI SENTIMEN

<i>Tweet</i>	Hasil Uji Sentimen
Dampak pandemi ini menjadi salah satu sorotan untuk usia >50 tahun yang tidak boleh bekerja di lapangan dan meningkatnya jumlah pengangguran. #UUCiptaKerja #OmnibusLaw	Negatif
Omnibus Law #UUCiptaKerja mendorong UMKM berinovasi digital.	Positif

V. KESIMPULAN

Kesimpulan yang dapat diberikan oleh penulis dan dapat diambil dari penelitian ini adalah sebagai berikut:

1. Pada analisis sentimen mengenai opini publik terhadap undang-undang cipta kerja pada media sosial twitter dengan mengimplementasikan metode *Naive Bayes Classifier*, dataset yang telah menyelesaikan tahapan preprocessing selanjutnya akan menghitung nilai bobot kata dengan mengimplementasikan algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)* yang selanjutnya akan diproses dengan metode *Naive Bayes Classifier* sehingga mendapatkan nilai dari *prior probability* dan juga nilai dari *conditional probability*. Selanjutnya akan diproses kedalam tahap testing. Dimana untuk mendapatkan hasil testing terbaik harus mencari nilai probabilitas tertinggi.
2. Implementasi metode *Naive Bayes Classifier* akan sangat optimal jika *data training* yang dipakai memiliki jumlah dataset tweet yang banyak. Implementasi Algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)* pada proses pembobotan kata yang dilanjutkan dengan implementasi metode *Naive Bayes Classifier* bekerja sangat baik dalam mengklasifikasikan teks dengan dataset *tweet* pada *twitter* dan juga mendapatkan sebuah hasil akurasi yang tinggi. Dalam penelitian ini penggunaan dataset yang dipakai akan dibagi menjadi *data training* dan *data testing* dengan perbandingan sebesar 60%:40%, 70%:30%, 80%:20%, 90%:10% dan didapatkan nilai akurasi tertinggi sebesar 94% dengan *data training* yang dipakai sebesar 90%, *data testing* sebesar 10%.
3. Berdasarkan hasil penelitian yang telah didapatkan pada tahap uji sentimen, diketahui bahwa hasil dari analisis sentimen opini publik terhadap undang-undang cipta kerja pada *twitter*, hasil uji sentimen menunjukkan hasil sentimen negatif didapatkan jika *tweet* yang diuji pada tahap uji sentimen mengandung kata yang bermakna negatif, penolakan, dan juga kontra, sedangkan hasil uji sentimen menunjukkan hasil sentimen positif didapatkan jika *tweet*

yang diuji pada tahap uji sentiment mengandung kata yang bermakna positif, pernyataan setuju dan juga mendukung.

UCAPAN TERIMA KASIH

Paper ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana (S1) pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Jember. Penulis mengucapkan terima kasih atas terbitnya naskah ini pada Seminar Nasional Sains Teknologi dan Inovasi Indonesia 2021 sebagai bagian kolaborasi/kerjasama penelitian antara Universitas Jember dengan Akademi Angkatan Udara. Ucapan serupa juga penulis sampaikan kepada pihak Twitter yang telah memberikan kemudahan dalam pengumpulan data dan seluruh pihak yang tidak bisa disebutkan satu per satu sehingga paper ini dapat diselesaikan.

REFERENSI

- [1] Mukaromah, Vina Fadhotul. 2020. "Ini Pro Kontra Yang Muncul Setelah Omnibus Law UU Cipta Kerja Disahkan Halaman All - Kompas.Com."2020. [serial on line] <https://www.kompas.com/tren/read/2020/10/07/143101665/ini-pro-kontra-yang-muncul-setelah-omnibus-law-uu-cipta-kerja-disahkan?page=all> [15 Oktober 2020].
- [2] Syam, Muh. Taufiq. 2020. "OPINI: Deviasi Informasi Media Sosial Di Masa Pandemi | IAIN PAREPARE." 2020. [serial on line] <https://www.iainpare.ac.id/opini-deviasi-informasi-media-sosial-di-masa-pandemi/> [16 Oktober 2020]
- [3] Karina, Fera dkk. 2018, Memaksimalkan Penggunaan Media Sosial Dalam Lembaga Pemerintah, Jakarta: Direktorat Jenderal Informasi dan Komunikasi Publik, Kementerian Komunikasi dan Informatika.
- [4] Suprpto Arifin, Hadi, Weny Widyawati, and Dan Taty Hernawaty. 2017. "Freedom of Expression di Media sosial Bagi Remaja Secara Kreatif dan Bertanggung Jawab: Bagi Siswa SMA AL-ma'soem Rancaek DAN SMA Muhammadiyah Pangandaran." *Jurnal Pengabdian Kepada Masyarakat*. 1(5): 332-337.
- [5] Riski, Wahyu Nova, and Yohanes Thianika Budiarsa. 2020. "Online Political Communication : An Overtime Analysis of Online Political Talk on Twitter During the 2019 Indonesian Presidential Election Campaign." *Jurnal Komunikasi Dan Media* 01 (01): 1–17. <http://journal.unika.ac.id/index.php/jkm/article/view/2845/pdf>.
- [6] Emeraldien, Fikry Zahria, Rifan Jefri Sunarsono, & Ronggo Alit. 2019. "Twitter Sebagai Platform Komunikasi Politik Di Indonesia" 14 (1): 21–30.
- [7] Rozi, Imam, Sholeh Pramono, and Erfan Dahlan. 2012. "Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi." *Jurnal EECIS* 6 (1): 37–43.
- [8] Nurdiansyah, Yanuar, Saiful Bukhori, & Rahmad Hidayat. 2018. "Sentiment Analysis System for Movie Review in Bahasa Indonesia Using Naive Bayes Classifier Method." *Journal of Physics: Conference Series* 1008 (1). <https://doi.org/10.1088/1742-6596/1008/1/012011>.
- [9] Pang, B., & Lee, L.(2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 1 (2): 91–233. <https://doi.org/10.1561/1500000001>.
- [10] Sundara, Tri A, Sherly Ekaputri Amas, and Sotar. 2020. "Naïve Bayes Classifier Untuk Analisis Sentimen Isu Radikalisme." *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)* 4 (1): 93–98.
- [11] Nurjanah, Winda Estu, Rizal Setya Perdana, & Mochammad Ali Fauzi. 2017. "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor Dan Pembobotan Jumlah Retweet." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)* Universitas Brawijaya 1 (12): 1750–57.
- [12] Lestari, Agnes Rossi Trisna, Rizal Setya Perdana, & M Ali Fauzi. 2017. "Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes Dan Pembobotan Emoji." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1 (12): 1718–24. <http://j-ptiik.ub.ac.id>.
- [13] Lilya Susanti. 2016, Metode Penelitian, Malang: Departemen Riset Teknologi dan Pendidikan Tinggi Universitas Brawijaya.