

ANALISIS SENTIMEN SEPUTAR UU ITE MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Yoga Vikriansyah Wijaya^{1*}, Adhitia Erfina², Cecep Warman³

^{1,2,3}Sistem Informasi, Universitas Nusa Putra

^{1,2,3}Jl. Raya Cibatu Cisaat No.21, Cibolang Kaler, Sukabumi Regency, Jawa Barat 43155

*Email Corresponding Author: yoga.vikriansyah_si17@nusaputra.ac.id

Abstrak

Analisis Sentimen cuitan twitter dari masyarakat Indonesia dapat dijadikan sebagai salah satu parameter untuk menjadi penunjang bagi pemerintah dalam mengevaluasi pengambilan keputusan dan kebijakan di masa yang akan datang. Penelitian ini bertujuan untuk mengetahui sentimen dari cuitan masyarakat Indonesia di twitter seputar Undang-Undang Informasi dan Transaksi Elektronik. Bahan data yang digunakan dalam penelitian ini menggunakan *query* Undang-Undang Informasi dan Transaksi Elektronik, Ujaran Kebencian, pencemaran nama baik, Penipuan *Online*, dan Pencurian data. Pengujian dilakukan dengan perhitungan akurasi, *precision*, *recall* dan *F1-score*, dengan menggunakan variatif data latih dan data uji. Hasil akurasi tertinggi didapatkan dari komposisi data latih 90% dan data uji 10% dengan nilai akurasi 84% dengan rata-rata *precision* 84%, *recall* 65%, *f1-score* 71% tiap kelas sentimen.

Kata Kunci: Analisis Sentimen, Algoritma Support Vector Machine, Cuitan Masyarakat

Abstract

Sentiment analysis of twitter tweets from the Indonesian people can be used as one of the parameters to be a support for the government in evaluating decision making and policies in the future. This study aims to find out the sentiments of Indonesian people's tweets on Twitter about the Information and Electronic Transaction Law. The data material used in this study uses a query on the Information and Electronic Transaction Law, Hate Speech, Defamation, Online Fraud, and Data Theft. The test is carried out by calculating accuracy, precision, recall and F1-score, using a variety of training data and test data. The highest accuracy results were obtained from the composition of 90% training data and 10% test data with an accuracy value of 84% with an average precision of 84%, recall 65%, f1-score 71% for each sentiment class.

Keywords: Sentiment Analysis, Support Vector Machine Algorithm, Community Tweet

1. Pendahuluan

Perkembangan teknologi Komunikasi yang begitu cepat memberikan sesuatu hal yang positif memudahkan kita dalam mendapatkan sebuah informasi atau menyampaikan suatu informasi melalui website atau media sosial, selain memberikan hal positif ternyata masih ada dampak negatif yang bisa terjadi dalam perkembangan teknologi komunikasi yaitu tindakan ujaran kebencian, pencurian data, penipuan transaksi *online* dan berita *hoax*.

Untuk mengatur kegiatan media komunikasi teknologi Undang-undang Informasi dan Transaksi Elektronik disahkan dan diundangkan pada tahun 2008 di Era Pemerintahan Presiden Susilo Bambang Yudhoyono dengan maksud memberikan rasa aman, keadilan dan kepastian hukum bagi pengguna dan penyelenggara teknologi informasi dan Transaksi Elektronik. Pada tahun 2016 UU ITE direvisi karena banyak perdebatan dari masyarakat dan pemahaman multitafsir terkait UU ITE. Menurut Samuel, revisi UU ITE berusaha membuat internet menjadi bermanfaat sebesar-besarnya bagi masyarakat sesuai dengan Undang-Undang Dasar. Selanjutnya, dengan revisi ini pemerintah berupaya memastikan kehadiran TIK bakal memperkuat integritas sosial dan mencerdaskan masyarakat, serta memajukan ekonomi digital berbasis kreativitas[1].

Saat ini kembali hangat diperbincangkan oleh masyarakat Indonesia, tokoh masyarakat atau para pejabat ketika pernyataan Presiden Republik Indonesia pada tanggal 15 februari 2021

yang intinya berisi, “Jika Undang-undang ITE tidak bisa memberikan rasa keadilan, Undang-undang ini akan direvisi”. Hal ini menyebabkan muncul berbagai pendapat oleh masyarakat dan memunculkan pro dan kontra mengenai wacana revisi UU ITE di media sosial. Media sosial yang digunakan dalam penelitian ini adalah twitter. Berdasarkan data dari *website* (<https://databoks.katadata.co.id/>) Twitter berada di peringkat 5 *platform* media sosial yang sering digunakan pada tahun 2020[3].

Pengaduan yang disampaikan masyarakat merupakan opini yang berharga dan salah satu alat evaluasi bagi pemerintah. Cuitan masyarakat seputar UU ITE, memiliki tatanan kalimat yang berbeda. Untuk itu diperlukan filterisasi keluhan dengan analisis sentimen. Manfaat analisis sentimen sangat penting untuk mengetahui sejauh mana data pengaduan masyarakat mengenai revisi undang-undang informasi dan transaksi elektronik digunakan sebagai alat untuk mengetahui pendapat dan reaksi masyarakat. Salah satu cara untuk mendapatkan informasi dari sekumpulan data dalam bentuk teks yaitu dengan membagi teks menjadi beberapa kategori yang telah ditentukan berdasarkan data yang kita miliki. Kinerja algoritma untuk klasifikasi teks sangat dipengaruhi oleh tipe data dan pemilihan fitur, dan beberapa teknik untuk mengekstraksi fitur bobot kata, seperti Word2Vec dan TF-IDF.

Dalam penelitian sebelumnya yang dilakukan oleh Sri Watmah, Suryanto dan Martias dengan judul Komparasi Metode K-NN, *Support Vector Machine*, Dan *Random Forest* Pada *E-Commerce* Shopee pada ulasan di google *playstore*, mengenai perbandingan beberapa algoritma dalam klasifikasi teks. Hasil tertinggi didapatkan dengan menggunakan algoritma SVM dengan hasil akurasi 89,4% presisi 89,5% dan *recall* 89,7% sedangkan untuk algoritma KNN hasil akurasi 89,0%, presisi 89,7% dan *recall* 87,5%[4]. Berdasarkan hasil akurasi tersebut maka dapat dibuktikan SVM memiliki performa yang stabil dan efektif dibandingkan dengan algoritma lainnya dalam melakukan klasifikasi teks.

Sebagai solusi untuk permasalahan, maka dibutuhkan penelitian yang komprehensif mengenai Analisis Sentimen masyarakat Indonesia seputar UU ITE. Kinerja dari sebuah algoritma klasifikasi dipengaruhi dari jenis data dan fiturnya, maka dari itu data set berupa teks yang akan diolah harus melalui tahapan *Text preprocessing* seperti *case folding*, *stemming*, *tokenizing*, *Text Normalization* serta *stopwords* dan juga pembobotan kata menggunakan TF-IDF lalu setelah itu data akan masuk tahapan selanjutnya yaitu tahapan klasifikasi menggunakan algoritma *Support vector Machine* kernel *linear* yang dapat mengklasifikasikan data untuk mendapatkan hasil kesimpulan.

2. Tinjauan Pustaka

2.1 Penelitian terkait

Auliya Rahman Isnain, Adam Indra Sakti, Debby Alita, Nurman Satya Marga penelitian pada tahun 2021 “Sentimen Analisis terhadap kebijakan *Lockdown* pemerintah Jakarta menggunakan Algoritma SVM”. Penelitian ini bertujuan untuk mengetahui bagaimana sentimen publik terhadap kebijakan yang akan dilakukan pemerintah mengenai kebijakan *lockdown* menggunakan algoritma *Support Vector Machine* dengan ekstraksi fitur tf-idf. Menghasilkan nilai akurasi sebesar 74%, *precision* sebesar 75%, *recall* sebesar 92% dan F1-Score sebesar 83%[5].

Faisal Rahutomo, Pramana Yoga Saputra, Miftahul Agramas Fidyawan “Implementasi Twitter Sentiment Analysis Untuk *Review* Film Menggunakan Algoritma *Support Vector Machine*” penelitian pada tahun 2018. Algoritma yang digunakan dalam klasifikasi adalah Support Vector Machine. Hasil akurasi klasifikasi algoritma Support Vector Machine menggunakan 60, 70, 80, dan 90 persen data latih rata-ratanya adalah 76,06 persen, 76,83 persen, 81,07 persen, dan 83,3 persen. Nilai *precision* positif memiliki rata-rata sebesar 79,97 persen, 78,71 persen, 84,02 persen, dan 85,54 persen. Nilai *precision* negatif memiliki rata-rata sebesar 81,73 persen, 87,41 persen, 87,37 persen, dan 93,61 persen. Nilai *precision* netral memiliki rata-rata sebesar 67,13 persen, 69,47 persen, 74,08 persen, dan 74,14 persen[6].

Pada tahun 2021 Penelitian yang dilakukan Primandani Arsi, Retno Waluyo yang berjudul Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine (Svm)*. Permasalahan yang ramai diperbincangkan di *Twitter* adalah terkait isu pemindahan ibu kota Indonesia. Hal ini menyebabkan munculnya fenomena perdebatan khususnya di *Twitter* yang sebenarnya menunjukkan perhatian kolektif mengenai wacana publik tersebut. Berdasarkan hasil pengujian yang dilakukan terhadap tweets sentimen pemindahan ibu kota dari media sosial twitter sebanyak 1.236 *tweets* (404 positif dan 832 negatif)

menggunakan SVM diperoleh akurasi =96,68%, *precision*=95.82%, *recall*=94.04% dan *AUC* = 0,979%[7].

Dalam Penelitian yang dilakukan Ragil Dimas Himawan dan eliyani yang berjudul Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. Pemerintah menggunakan Twitter sebagai saluran untuk berinteraksi dengan masyarakat, dan pandemi COVID-19 ada di akun Twitter resmi Pemprov DKI Jakarta. Pemprov DKI Jakarta, dengan *username* twitter @dkijakarta, sehingga masyarakat dapat mengungkapkan pandangan mereka mengajukan pertanyaan mengenai layanan publik dan tentang kebijakan pemerintah, Data cuitan dikategorikan menjadi sentimen negatif, netral, dan positif. Gunakan TF-IDF *Vectorizer* untuk pembobotan kata dan gunakan beberapa metode untuk klasifikasi yaitu Algoritma *Naïve Bayes* akurasi 75,22%, *random forest classifier* dengan hasil akurasi 75,81%, dan algoritma *Support Vector Machine* adalah 77,58%. Dalam proses analisis sentimen pada tweet, persentase hasil negatif, netral, dan positif berturut-turut adalah 8,8%, 83,6%, dan 7,6%[8].

Perbedaan dari penelitian-penelitian yang sudah dilakukan sebelumnya, yaitu penelitian mengenai Analisis Sentimen Seputar UU ITE Menggunakan Algoritma *Support Vector Machine* menggunakan Bahasa pemrograman Python 3.7 untuk proses pengolahannya. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Support Vector Machine* kernel *linear* terhadap dataset seputar UU ITE mengenai ujaran kebencian, pencemaran nama baik, penipuan *online*, dan pencurian data dengan dataset berjumlah 3000. Penelitian ini diharapkan dapat dijadikan sebagai salah satu parameter untuk menjadi penunjang bagi pemerintah dalam mengevaluasi pengambilan keputusan dan kebijakan di masa yang akan datang dan memberikan gambaran hasil pelabelan sentimen.

2.2 Landasan Teori

2.2.1 Undang-undang Informasi dan Transaksi Elektronik

Undang-undang informasi dan Transaksi Elektronik adalah undang-undang yang mengatur tentang informasi Elektronik dan Transaksi Elektronik. Informasi Elektronik adalah sebagai suatu kumpulan data elektronik dapat berupa gambar, peta, rancangan, foto, *electronic data interchange* (EDI), pesan elektronik seperti e-mail, telegram. Sedangkan Transaksi Elektronik adalah suatu perbuatan hukum yang dilakukan menggunakan komputer, jaringan komputer, atau media elektronik lainnya[9].

2.2.2 Analisis Sentimen

Analisis Sentimen adalah teknik untuk mengidentifikasi pendapat orang, emosi terhadap situasi dan sikap. Analisis sentimen berfokus pada opini-opini yang mengekspresikan atau mengungkapkan sentimen positif netral dan negatif berdasarkan kata-kata yang digunakan dalam kalimat pendapat masyarakat[10].

2.2.3 Python

Python adalah bahasa pemrograman tingkat tinggi dan dapat digunakan untuk memproses data, membangun situs web, dan jaringan. Istilah bahasa pemrograman tingkat tinggi mengacu pada tingkat kedekatan program dengan kode listrik yang digunakan oleh komputer. Suatu bahasa pemrograman disebut bahasa pemrograman tingkat tinggi karena perintah atau kode program yang digunakan mirip dengan bahasa manusia.

Python terinspirasi oleh bahasa pemrograman ABC. Hingga saat ini Guido masih menjadi penulis utama untuk python, meskipun bersifat *open source* sehingga ribuan orang juga berkontribusi dalam mengembangkannya[11].

2.2.4 Pembobotan Kata

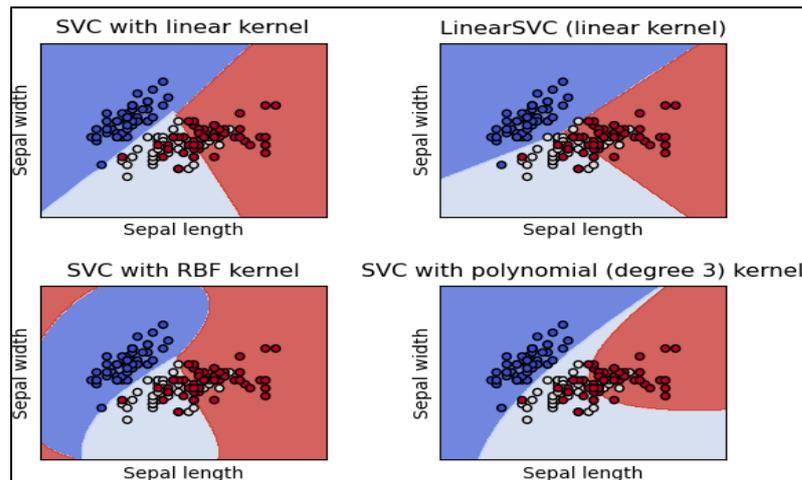
Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah proses pengubahan data dalam data teks menjadi data numerik dan pembobotan setiap kata atau fitur. TF-IDF adalah ukuran statistik yang digunakan untuk mengevaluasi pentingnya kata-kata dalam sebuah dokumen. TF adalah frekuensi kata dalam dokumen tertentu, yang menunjukkan pentingnya kata dalam setiap dokumen. DF adalah frekuensi dokumen yang mengandung kata, menunjukkan seberapa umum kata tersebut. IDF adalah kebalikan dari nilai DF. Hasil dari penggunaan TF-IDF untuk bobot kata adalah hasil kali TF dikalikan dengan IDF. Jika kata lebih sering muncul dalam satu dokumen, bobot kata lebih besar, dan jika muncul di banyak dokumen, bobotnya lebih kecil[12].

2.2.5 Support Vector Machine

Memahami bahwa algoritma *support vector machine* (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi linier dalam fitur berdimensi tinggi dan menggunakan algoritma pembelajaran berdasarkan teori optimasi untuk pelatihan[13].

Pengertian algoritma *Support Vector Machine* (SVM) adalah sistem pembelajaran spasial hipotetis berupa fungsi linier dalam fitur multidimensi dan dilatih menggunakan algoritma pembelajaran berbasis teori optimasi[13]. Prinsip dasar SVM adalah pengklasifikasi linier, dan dikembangkan untuk bekerja dengan masalah nonlinier. dengan memasukkan konsep trik kernel ke dalam ruang kerja dimensional [14].

Metode kernel dalam Algoritma *Support Vector Machine* adalah cara untuk melakukan proses klasifikasi terhadap data yang tidak terklasifikasi secara *linear*[15].

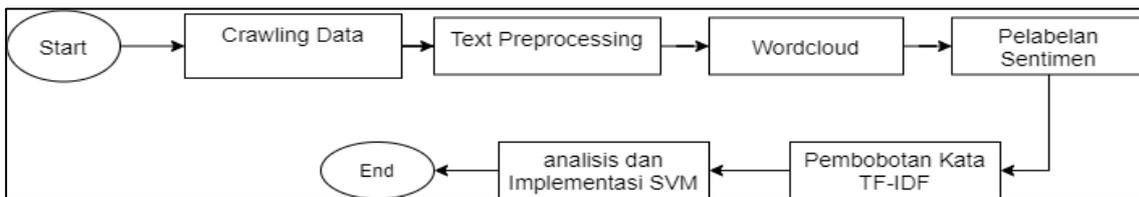


Gambar 1. Klasifikasi Multi-Kelas SVM

(Sumber: https://scikit-learn.org/stable/auto_examples/svm/plot_iris_svc.html)

Pada Gambar 2.2 terlihat bahwa model klasifikasi data menggunakan metode SVC memiliki beberapa kernel yaitu kernel linier, kernel RBF dan kernel polinomial. SVC adalah perintah yang paling umum digunakan dalam algoritma SVM. Algoritma SVM pada SVC membagi garis tengah hyperplane menjadi linier dua dimensi, tanpa menambahkan kernel lain. SVC menggunakan kernel linear dianggap cukup untuk mendeteksi keakuratan data.

3. Metodologi

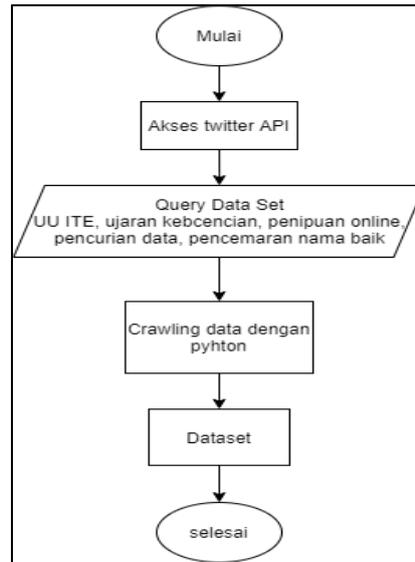


Gambar 2. Diagram Kerja

1) Crawling Data

Teknik Pengumpulan data pada penelitian ini menggunakan *crawling* data. Data yang diambil adalah data cuitan yang terdapat di twitter menggunakan koneksi Twitter API, untuk akses memerlukan hak akses berupa API key, API secret key, access token, dan access token secret[16]. Data yang dikumpulkan berupa teks, Proses *Crawling*/Pengumpulan data dalam penelitian ini menggunakan Bahasa pemrograman python 3.7 dengan bantuan *library* tweepy dari python. Berikut alur *crawling* data.

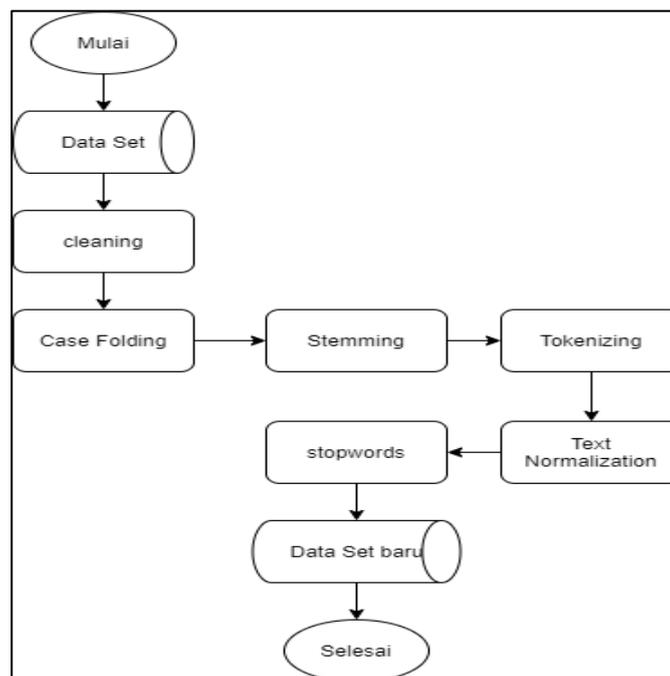
Data yang diambil adalah bahasa indonesia dengan menggunakan kata kunci UU ITE, ujaran kebencian, Penipuan *Online*, pencemaran nama baik dan Pencurian Data.



Gambar 3. Alur crawling Data

2) **Text Preprocessing**

Setelah melakukan pengumpulan data dataset belum siap untuk digunakan maka dari itu dilakukan *Text Preprocessing* pada dataset agar menghasilkan data yang siap digunakan dalam proses selanjutnya. Berikut alir kerja *Text Preprocessing*.



Gambar 4. Alur Text Preprocessing

Dari gambar 4 diatas tahap *text preprocessing* pertama adalah *cleaning* yaitu untuk membersihkan data teks dari link, nomor, tanda baca, RT dan emoji. Kemudian dilakukan *case folding* untuk mengubah semua huruf dalam bentuk huruf kecil, setelah itu melakukan proses *stemming* dengan tujuan mentransformasi kata menjadi kata baku, lalu melakukan *tokenizing* untuk memecah kalimat menjadi potongan kata-kata, lalu setelah itu melakukan *text normalization* yaitu mengubah kata yang disingkat menjadi kata baku. Dan yang terakhir melakukan proses *stopwords* untuk menghilangkan kata-kata yang tidak terpakai atau tidak informatif pada kalimat.

3) Word Cloud

Lalu setelah melalui tahapan *text preprocessing* selanjutnya adalah *Word cloud*. *Word cloud* yaitu untuk visualisasi dari sebuah dokumen terhadap kata-kata yang sering muncul pada sebuah dokumen. Frekuensi dari kata yang sering muncul ditunjukkan melalui ukuran suatu teks tersebut. Apabila ukuran teks semakin besar menunjukkan frekuensi kata tersebut muncul dalam dokumen[17].

4) Pelabelan Sentimen

Sesudah melalui tahapan *text preprocessing* langkah selanjutnya analisis sentimen untuk pelabelan data sentimen. Pelabelan sentimen adalah salah satu cara untuk mengetahui pendapat seseorang atau terhadap suatu isu. Pelabelan sentimen pada penelitian ini dibagi menjadi 3 yaitu Sentimen Negatif, netral dan positif. Kalimat tweet yang memiliki nilai > 0 diklasifikasikan dalam kelas positif, kalimat tweet yang memiliki $= 0$ diklasifikasikan kelas netral, dan kalimat tweet yang memiliki nilai < 0 diklasifikasikan kedalam kelas negatif.

Dalam proses pelabelan sentimen, peneliti membuat kamus kata positif dan negatif bahasa Indonesia dengan membuat suatu sistem menggunakan Bahasa program python 3.7 agar bisa melakukan pelabelan kalimat yang mengandung kata positif, negatif dan netral yang sesuai terdapat pada kamus tersebut,

5) Pembobotan TF-IDF

Pada penelitian ini pembobotan kata menggunakan algoritma TF-IDF menggunakan *TfidfVectorizer* library *scikit-learn* dari python untuk memudahkan implementasi TF-IDF mengubah data teks menjadi bentuk vektor.

6) Analisis dan Implementasi SVM

Setelah melalui *text preprocessing*, pelabelan sentimen, dan Pembobotan kata, proses selanjutnya melakukan klasifikasi dengan algoritma *Support Vector Machine* kernel *linear*, Alasan menggunakan kernel ini adalah karena kernel ini paling sederhana yang umum digunakan dan memiliki waktu pemrosesan yang lebih sedikit.

Proses klasifikasi yaitu mempelajari pola dari data dengan menggunakan data latih. Data latih yang didalamnya terdapat cuitan negatif, positif, dan netral digunakan dalam proses algoritma SVM dalam mempelajari pola data berdasarkan kategori sentimen masing-masing pada data. Hasil pembelajaran dengan algoritma SVM lalu dapat diuji menggunakan data uji, sehingga akurasi dapat diukur saat memprediksi kategori dalam data baru.

Pada penelitian ini menggunakan evaluasi *confusion matrix multi class 3x3*, untuk mendapatkan hasil akurasi, *precision*, *recall*, dan *F1-score*. Untuk Implementasi algoritma *Support Vector Machine* menggunakan metode kernel *linear*, dengan Bahasa pemrograman python dan dibantu menggunakan library *Scikit-Learn*.

4. Hasil dan Pembahasan

4.1 Crawling Data

Crawling data twitter ini dilaksanakan kurang lebih selama 1 bulan pada tanggal 1 maret sampai 7 april, dengan jumlah 3000, dalam seminggu dilakukan 5 kali *crawling* data per *query* 150 data yaitu *query*, uu ite, pencemaran nama baik, pencurian data, ujaran kebencian, dan penipuan *online*. Hasil *crawling* data dapat dilihat pada gambar 5.

	user	tanggal	tweet
0	Caesar Dcandra	2021-03-04 01:59:00	'@SKakilima "@CNNIndonesia smentara (sementah...
1	Savvy?	2021-03-04 00:59:00	blokir uuite when'
2	Jacob Elordi	2021-03-03 16:47:00	'@wafxr tiati kena uuite'
3	acehsatu media	2021-03-03 16:41:00	#KisahPilu #IbudanBayiDipenjara #UUIE Kasus ...
4	detikcom	2021-03-03 14:59:00	Mereka menyampaikan pengalamannya berurusan de...
...
2995	Si Ganjang Ulu	2021-03-16 15:11:00	TIPS MENCEGAH PENCURIAN DATA KARTU KREDIT n \n...
2996	BukaMata	2021-03-16 14:28:00	Hacking dan cracking sebenarnya sama saja, bed...
2997	POLSEKWUNGU. ID	2021-03-16 13:05:00	Tips mencegah pencurian data kartu kredit n \n...
2998	DetikFinance	2021-03-16 13:02:00	#mostpop Aplikasi ini disebut-sebut bisa mengg...
2999	Pati Kita	2021-03-16 10:39:00	RT "@ResorPati: TIPS MENCEGAH PENCURIAN DATA ...

3000 rows x 3 columns

Gambar 4. Hasil Crawling

4.2 Text Preprocessing

Setelah mendapatkan hasil *crawling* data dan dijadikan dataset, tahap selanjutnya adalah *text preprocessing* data. Berikut contoh *text preprocessing* pada tabel 1.

Tabel 1. Contoh Text Preprocessing

Tahapan Text Preprocessing	Sebelum	Sesudah
<i>Cleaning</i>	Jika konten yang berpotensi melanggar UU ITE belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali #PolisiVirtual #UUIE #terasnegeriku	Jika konten yang berpotensi melanggar UU ITE belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali PolisiVirtual UUIE terasnegeriku
<i>Case Folding</i>	Jika konten yang berpotensi melanggar UU ITE belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali PolisiVirtual UUIE terasnegeriku mengirimkan pesan peringatan kembali PolisiVirtual UUIE terasnegeriku	jika konten yang berpotensi melanggar uu ite belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali polisivirtual uuite terasnegeriku
<i>Stemming</i>	jika konten yang berpotensi melanggar uu ite belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali polisivirtual uuite terasnegeriku	jika konten yang potensi langgar uu ite belum turun, polisi virtual akan kirim pesan ingat kembali polisivirtual uuite terasnegeriku
<i>Tokenizing</i>	jika konten yang berpotensi melanggar uu ite belum diturunkan, polisi virtual akan mengirimkan pesan peringatan kembali polisivirtual uuite terasnegeriku	[jika, konten, yang, potensi, langgar, uu, ite, belum, turun, polisi, virtual, akan, kirim, pesan, ingat, kembali, polisivirtual, ,uuite, terasnegeriku]
<i>Text Normalization</i>	Tiati kena uuite	Hati hati kena uu ite
<i>Stopwords</i>	Giliran Ravio Patra hingga Nikita Mirzani Diminta Masukan terkait Revisi UU ITE NikitaMirzani UUIE	giliran ravio patra hingga nikita mirzani diminta masukan terkait revisi uu ite nikitamirzani uuite

Tabel 3. Jumlah data pada kelas sentimen

Sentimen	Jumlah
Negatif	2223
Netral	438
Positif	339

Dari total cuitan di dataset sebanyak 3000, hasil pelabelan kelas sentimen berdasarkan tabel 4.2 menunjukkan bahwa cuitan negatif memiliki frekuensi yang lebih tinggi dibandingkan cuitan netral dan positif. Perhitungan Persentase klasifikasi sentimen diperoleh berdasarkan rumus berikut:

$$\frac{\text{Jumlah komentar yang terklasifikasi}}{\text{jumlah seluruh komentar}} \times 100\%$$

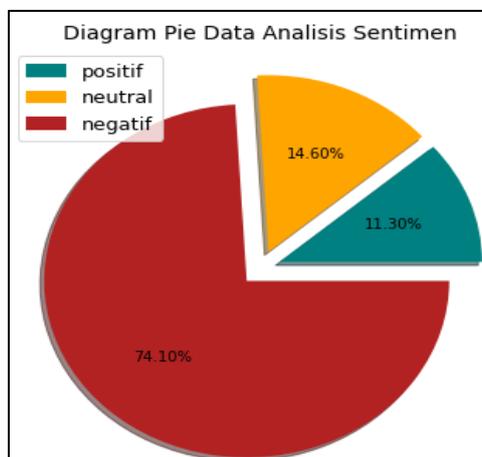
Berdasarkan dataset cuitan twitter mengenai UU ITE diperoleh sebanyak 3000 yang terdiri dari 2223 negatif, 438 netral dan 339 positif. Kemudian dihitung persentase sentimen sesuai rumus diatas. Perhitungan persentase diperoleh sebanyak berikut:

$$\text{Negatif} = \frac{2.223}{3000} \times 100\% = 74.10\%$$

$$\text{Netral} = \frac{438}{3000} \times 100\% = 14.60\%$$

$$\text{Positif} = \frac{339}{3000} \times 100\% = 11.30\%$$

Persentase klasifikasi sentimen pada dataset dapat dilihat pada gambar 7.



Gambar 7. Bar Diagram Klasifikasi Sentimen

4.5 Pembobotan Kata

Pembobotan kata yang digunakan untuk penelitian ini yaitu TF-IDF (*term Frequency, Inverse Document Frequency*). Pembobotan kata dilakukan dengan menghitung banyaknya kata yang muncul dalam suatu dokumen. Pembobotan kata dilakukan setelah melewati tahap *text preprocessing*.

Pembobotan bertujuan untuk memberikan nilai kepada *term* dan nilai *term* akan dijadikan sebagai input untuk proses klasifikasi. Dengan menggunakan python dan bantuan *library* scikit learn *Class TfIdfVectorizer* mampu menghitung nilai TF-IDF:

```
# TF-IDF
#untuk pembobotan kata
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vect = TfidfVectorizer()
tfidf_vect.fit(df['Tweet'])
train_X_tfidf = tfidf_vect.transform(df_train['Tweet'])
test_X_tfidf = tfidf_vect.transform(df_test['Tweet'])
tfidf_vect
```

Gambar 8. Proses TF-IDF

Pada gambar 8 adalah proses TF-IDF *Vectorizer* adalah proses transformasi data dalam bentuk teks menjadi ke bentuk vector. Berikut contoh hasil pembobotan kata menggunakan TF-IDF dengan programan python.

Tabel 3. Contoh hasil Pembobotan Kata dengan TF-IDF di Python

Urutan dan kata	Hasil TF-IDF
(0, 5332)	0.15125827583407356
(0, 3954)	0.21122472887469904
(0, 3731)	0.23417918743086774
(1, 5030)	0.36822523317312017
(1, 5011)	0.37933300194623315
(1, 3846)	0.4138368832461843

Penjelasan angka 0 dalam kolom urutan dan kata seperti (0, 5332) dijelaskan bahwa angka 0 melambangkan kalimat/tweet pertama dan angka 5332 melambangkan kata ke 5332 pada urutan kata di penggabungan, dan 0.15125827583407356 di kolom hasil TF-IDF melambangkan nilai dari pembobotan kata TF-IDF.

4.6 Klasifikasi Dengan *Support Vector Machine*

Pada penelitian ini, dataset yang diperoleh dari *tweets* seputar UU ITE di twitter. Didapatkan data sebanyak 3000 setelah melalui tahapan *text preprocessing*, pembobotan *term frequency-inverse document frequency* (tf-idf) dan pelabelan, kemudian data tersebut dibagi menjadi 2 untuk data latih dan data uji, algoritma yang digunakan yaitu *Support Vector Machine* kernel *linear*. Persentase akurasi dihasilkan berdasarkan nilai dari label sentimen, setiap label memiliki nilai berbobot Berkas yang dihitung. Pada pengujian di penelitian ini menggunakan evaluasi *confusion matrix*, untuk mendapatkan hasil berupa nilai akurasi, *precision*, *recall*, dan F1 score dari tiap kelas sentimen dan rata-rata dari *precision*, *recall* dan F1-Score. Berikut rumus *confusion matrix multiple class 3x3* untuk mencari hasil akurasi, *precision*, *recall* dan F1-score.

Tabel 4. Pembagian Data latih dan data Uji

Perbandingan Data		Jumlah Data	
Data Latih	Data Uji	Data Latih	Data Uji
90%	10%	2700	300
80%	20%	2400	600
70%	30%	2100	900
60%	40%	1800	1200
50%	50%	1500	1500
40%	60%	1200	1800
30%	70%	900	2100
20%	80%	600	2400
10%	90%	2700	300

Pada tabel 4 dijelaskan pada pengujian dilakukan sebanyak 9 kali dengan komposisi data latih dan data uji yang variatif untuk digunakan klasifikasi data. Berikut *script* pemanggilan algoritma

algoritma *Support Vector Machine* kernel *linear* dan proses klasifikasi dengan menggunakan *confusion matrix* pada gambar 8.

```

from sklearn.svm import SVC

model = SVC(kernel='linear')
model.fit(train_X_tfidf,train_Y)

predictions_SVM = model.predict(test_X_tfidf)
test_prediction = pd.DataFrame()
test_prediction['Tweet'] = test_X
test_prediction['Label'] = predictions_SVM

# Accuracy, Precision, Recall, f1-score

from sklearn.metrics import f1_score, recall_score, precision_score, confusion_matrix, accuracy_score, classification_report
print (confusion_matrix(test_Y, predictions_SVM))
print (classification_report(test_Y, predictions_SVM))
print ("nilai akurasi dengan SVM adalah :", accuracy_score(test_Y, predictions_SVM)*100)
print("SVM Recall Score :", recall_score(test_Y, predictions_SVM,average='macro')*100)
print("SVM Precision Score :", precision_score(test_Y, predictions_SVM,average='macro')*100)
print("SVM F1 Score :", f1_score(test_Y, predictions_SVM,average='macro')*100)
    
```

Gambar 9. Script proses klasifikasi SVM

Berikut adalah hasil klasifikasi dari beberapa tahap pengujian:

Tabel 5. Hasil Klasifikasi *Support Vector Machine*

Split Data		Akurasi	Kelas Sentimen	Precision	Recall	F1-score
Data Latih	Data Uji					
90%	10%	84%	Negatif	0.84	0.98	0.90
			Netral	0.73	0.39	0.51
			Positif	0.95	0.57	0.71
			Rata-Rata (%)	84%	65%	71%
80%	20%	82%	Negatif	0.83	0.98	0.90
			Netral	0.69	0.31	0.43
			Positif	0.91	0.61	0.73
			Rata-Rata (%)	81%	63%	69%
70%	30%	81%	Negatif	0.81	0.98	0.89
			Netral	0.77	0.28	0.42
			Positif	0.89	0.52	0.66
			Rata-Rata (%)	78%	60%	65%
60%	40%	80%	Negatif	0.80	0.98	0.88
			Netral	0.73	0.24	0.36
			Positif	0.86	0.47	0.61
			Rata-Rata (%)	80%	57%	62%
50%	50%	80%	Negatif	0.80	0.99	0.88
			Netral	0.72	0.21	0.33
			Positif	0.86	0.46	0.60
			Rata-Rata (%)	80%	55%	60%
40%	60%	80%	Negatif	0.79	0.99	0.88
			Netral	0.80	0.18	0.30
			Positif	0.89	0.40	0.55
			Rata-Rata (%)	79%	50%	58%
30%	70%	80%	Negatif	0.79	0.99	0.88
			Netral	0.78	0.17	0.28
			Positif	0.93	0.37	0.53
			Rata-Rata (%)	83%	51%	56%
			Negatif	0.77	1.00	0.87

Split Data		Akurasi	Kelas Sentimen	Precision	Recall	F1-score
Data Latih	Data Uji					
20%	80%	78%	Netral	0.84	0.12	0.21
			Positif	0.95	0.19	0.32
			Rata-Rata (%)	85%	44%	46%
10%	90%	76%	Negatif	0.76	1.00	0.86
			Netral	1.00	0.10	0.18
			Positif	0.97	0.09	0.17
			Rata-Rata (%)	91%	40%	40%

Berdasarkan hasil pengujian yang dilakukan untuk mencari nilai akurasi, dan nilai rata-rata dari *precision*, *recall* dan *F1-Score*, terdapat perbedaan nilai akurasi, *precision*, *recall* dan *f1-score*, pengujian yang dilakukan dengan jumlah data latih yang berbeda. Dari hasil pengujian tersebut didapatkan hasil yang terbesar dengan data latih 90% dan data uji 10% menghasilkan nilai akurasi yang tertinggi dengan nilai akurasi 84%, dengan rata-rata tiap kelas sentimen 84% *Precision*, 65% *Recall* dan 71% *F1-score*. menunjukkan bahwa pada penelitian ini semakin besar data latih dan semakin kecil data uji yang digunakan memberikan hasil akurasi yang lebih tinggi.

5. Kesimpulan

Berdasarkan hasil percobaan data teks yang diambil di twitter yang diolah menjadi analisis sentimen dengan menggunakan *tf-idf* dan algoritma *Support Vector Machine* kernel *linear* dapat diambil beberapa kesimpulan yaitu:

1. *TF-IDF* membantu Algoritma *Support Vector Machine* kernel *Linear* untuk melakukan pembobotan kata dalam analisis teks.
2. Dari 9 kali pengujian dengan menggunakan algoritma *Support Vector Machine* kernel *linear* dengan pembagian data variatif didapatkan hasil akurasi tertinggi 84%, dengan rata-rata tiap kelas sentimen *precision* 84%, *recall* 65%, *f1-score* 71% dari data latih 90% dan data uji 10%.
3. Proses *text preprocessing* sangat berpengaruh terhadap hasil klasifikasi untuk menghasilkan nilai akurasi yang tinggi.
4. Dalam penelitian ini menggunakan bahasa pemrograman Python 3.7, maka perlu adanya penggunaan aplikasi atau bahasa program lain yang bisa mengolah data untuk analisis sentimen dalam penelitian ini, untuk mempengaruhi dalam proses klasifikasi.

DAFTAR REFERENSI

- [1] Agustin Setyo Wardani, "Alasan Pemerintah Revisi UU ITE," 2016. <https://www.liputan6.com/tekno/read/2664146/alasan-pemerintah-revisi-uu-ite> (accessed Feb. 24, 2021).
- [2] CNN Indonesia, "Angin Kencang Revisi UU ITE dari Jokowi hingga Mahfud MD," 2021. <https://www.cnnindonesia.com/nasional/20210216075242-32-606680/angin-kencang-revisi-uu-ite-dari-jokowi-hingga-mahfud-md> (accessed Feb. 24, 2021).
- [3] Katadata.com, "10 Media Sosial yang Paling Sering Digunakan di Indonesia," 2020. <https://databoks.katadata.co.id/datapublish/2020/02/26/10-media-sosial-yang-paling-sering-digunakan-di-indonesia#> (accessed Feb. 21, 2021).
- [4] Shopee, P.E., Watmah, S. Komparasi Metode K-NN , Support Vector Machine , Dan Random Forest, 2021, 2(1): 15–21
- [5] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *Jurnal Data Mining Dan Sistem Informasi*, 2021, 2(1): 31-37.
- [6] Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A . Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*, 2018, 4(2): 93-93.
- [7] Arsi, P., & Waluyo, R. Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 2021, 8(1): 147-155.
- [8] Himawan, R. D., & Eliyani, E. Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 2021, 7(1): 58-63.
- [9] Admin DSLA, "UU ITE: Pasal-pasal dan Mereka yang Terjerat," 2020. <https://www.dslalawfirm.com/uu-ite/> (accessed Feb. 26, 2021).
- [10] A. Erfina and Y. H. Putra, "Irony Sentence Detection Techniques Using Fuzzy Historical Classifier," *IOP Conf. Ser. Mater. Sci. Eng.*, 2019, 662(6).
- [11] S. Falentino and A. Erfina, *BAHASA ULAR UNTUK PEMROGRAMAN PYTHON*. indonesia: Insan Cendekia Mandiri, 2020, 2020.
- [11] Melita, R. *Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta), 2018
- [12] Melita, R. *Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)* *J. Tek. Inform.*, 2018, 11(2): 149–164.
- [13] Ruliah, R. Klasifikasi Stadium Trophozoite, Schizonts, Gametocyter pada Sediaan Darah Plasmodium Falciparum dengan Pendekatan Support Vector Machine. *Progresif: Jurnal Ilmiah Komputer*, 2015, 6(2): 641-686
- [14] Wahyudin, F.A.I. Eneng Tita Tosida, *Teori dan Panduan Praktis Data Science dan Big Data*, I. indonesia: Lembaga Penelitian dan Pengabdian pada Masyarakat Universitas Pakuan, 2019.
- [15] Maisarah, M., Ayu, "SISTEM ANALISIS SENTIMEN PADA FANPAGE FACEBOOK KANDIDAT PRESIDEN 2019-2024," Universitas Sumatera Utara, 2020.
- [16] Sari, D. I., Wati, Y. F., & Widiastuti, W. W. ANALISIS SENTIMEN DAN KLASIFIKASI TWEETS BERBAHASA INDONESIA TERHADAP TRANSPORTASI UMUM MRT JAKARTA MENGGUNAKAN NAÏVE BAYES CLASSIFIER. *Jurnal Ilmiah Informatika Komputer*, 2020, 25(1): 64-75.

- [17] Kurniawan, T. "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine," Institut Teknologi Sepuluh Nopember, 2017.