

ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Primandani Arsi^{*1}, Retno Waluyo²

¹Informatika, Universitas Amikom Purwokerto, ²Sistem Informasi, Universitas Amikom Purwokerto
Email: ¹ukhti.prima@amikompurwokerto.ac.id, ²waluyo@amikompurwokerto.ac.id

(Naskah masuk: 21 Agustus 2020, diterima untuk diterbitkan: 02 Februari 2021)

Abstrak

Dewasa ini, media sosial berkembang pesat di internet, salah satu yang banyak digemari adalah Twitter. Berbagai topik ramai diperbincangkan di Twitter mulai dari ekonomi, politik, sosial, budaya, hukum dan lain-lain. Salah satu topik yang ramai diperbincangkan di Twitter adalah terkait isu pemindahan ibu kota Indonesia. Namun dibalik hal tersebut terdapat kontroversi dari pihak yang merasa pro dan kontra, masing-masing memiliki sudut pandang yang berbeda. Hal ini menyebabkan munculnya fenomena perdebatan khususnya di Twitter yang sebenarnya menunjukkan perhatian kolektif mengenai wacana publik tersebut. Analisis sentimen adalah proses mengekstraksi, memahami dan mengolah data berupa teks yang tidak terstruktur secara otomatis guna mendapatkan informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini. Dalam penerapan analisis sentimen menggunakan metode *machine learning* terdapat beberapa metode yang sering digunakan. Dalam penelitian ini di gunakan metode *Support Vector Machine* (SVM) untuk diterapkan pada *tweets* topik pemindahan ibu kota Indonesia untuk tujuan klasifikasi kelas sentimen pada media sosial *twitter*. Teknis klasifikasi dilakukan dengan cara mengklasifikasikan menjadi 2 kelas yakni positif dan negatif. 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.

Kata kunci: *twitter, ibu kota, analisis sentimen, support vectore machine (SVM).*

SENTIMENT ANALYSIS ON THE DISCUSSION OF RELOCATING INDONESIA'S CAPITAL CITY USING THE SUPPORT VECTOR MACHINE (SVM)

Abstract

Today, social media is growing fast on the internet. One of the most popular social media is Twitter. Many topics are discussed on Twitter such as economic, politic, social, culture, and law. One of the hot topics discussed on Twitter is the issue of relocating Indonesia's capital city. However, there is controversy from supporters and opponents. They have different views. This issue leads to a phenomenon of debate on Twitter that actually shows a collective concern about the public discourse. Sentiment analysis is a process of extracting, understanding and processing unstructured data to get sentiment information which is found in an opinion sentence. Application of sentiment analysis using machine learning methods shows that there are several methods that are often used. In this study, the Support Vector Machine (SVM) method is proposed to be applied to tweets on the topic of relocating Indonesia's capital city for sentiment classification on social media twitter. The classification technique is carried out into 2 classes, namely positive and negative. Based on testing on the sentiment of relocating Indonesia's capital city from social media twitter from 1,116 tweets (404 positive and 832 negative) using SVM obtained accuracy = 96.68%, precision = 95.82%, recall = 94.04% and AUC = 0.979.

Keyword : *twitter, capital city, sentiment analysis, support vectore machine (SVM).*

1. PENDAHULUAN

Dewasa ini, media sosial berkembang pesat di internet, salah satu yang banyak digemari adalah Twitter (Zulfa dan Winarko, 2017). Seperti media

sosial lainnya, Twitter merupakan platform media sosial yang memungkinkan penggunaanya untuk berinteraksi secara personal ataupun terbuka. Berbagai topik ramai diperbincangkan di Twitter mulai dari politik, ekonomi, sosial, budaya, dan hukum (Teran

dan Mancera, 2019) (Kusen dan Stembeck, 2018) (Öztürk dan Ayvaz, 2017). Melalui fitur *hashtag* para pengguna Twitter dapat mengetahui topik yang sedang dibahas secara *real-time*. Selain itu kata kunci pada Twitter dapat pula menjadi sumber perbincangan oleh pengguna. Tidak jarang sebelum suatu berita muncul sebagai *headline* di media *online* pembahasannya sudah terlebih dahulu dibahas di Twitter. Bahkan pembahasan di Twitter memicu suatu berita muncul (Priyatno dkk., 2019). Saat ini penggunaan *Twitter* khususnya di Indonesia telah memberikan dampak besar pada sudut pandang topik tertentu (Pandey, Rajpoot dan Saraswat, 2017).

Salah satu topik yang *trending* di Twitter adalah terkait isu pemindahan ibu kota Indonesia. Wacana tentang pemindahan ibu kota senantiasa muncul disetiap era presiden, namun wacana tersebut kembali *trending* setelah pemerintahan Presiden Jokowi membahas hal tersebut kembali (Septiana, 2018). Wacana pemindahan merupakan bagian dari kepentingan bangsa dan sebagai upaya menyongsong persaingan global (Toun, 2018). Namun dibalik hal tersebut terdapat kontroversi dari pihak yang merasa pro dan kontra, masing-masing memiliki sudut pandang yang berbeda. Hal ini menyebabkan munculnya fenomena perdebatan khususnya di Twitter yang sebenarnya menunjukkan perhatian kolektif mengenai wacana publik tersebut (Andriani, 2018). Kecenderungan pengguna Twitter dalam memposting konten dapat diketahui dengan cara analisis sentimen (Pratama, Andrean dan Nugroho, 2019). Dengan cara memisahkan opini berdasarkan kelas yakni positif dan negatif. Selain itu dilakukan teknik pengambilan kesimpulan terkait faktor apa yang sering dibahas dalam opini tersebut (Novantirani, Sabariah dan Effendy, 2015).

Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini (Brahimi, Touahria dan Tari, 2019). Analisis sentimen dilakukan guna menilai opini dan kecenderungan sebuah opini terhadap suatu topik baik negatif maupun positif (Rozi, Pramono dan Dahlan, 2012). Analisis sentimen dapat diterapkan pada opini semua bidang seperti ekonomi, politik, sosial dan hukum. Media sosial Twitter ini membuka jendela bagi para peneliti untuk mempelajari emosi, suasana hati, dan pendapat publik melalui analisis sentimen (Qiu, Lin dan Shuai, 2019).

Dalam penerapan analisis sentimen menggunakan metode *machine learning* terdapat beberapa metode yang sering digunakan seperti K-NN, *Naive Bayes*, *Random Forest* dan *Support Vector Machine (SVM)* (Zuhdi, Utami dan Raharjo, 2019) (Wang, Liu dan Ma, 2019) (Fauzi, 2018) (Thakur dan Deshpande, 2019). Salah satu penelitian yang memanfaatkan data media sosial Twitter dengan metode *machine learning* adalah Lukmana dkk, dalam penelitiannya digunakan data politik dari twitter

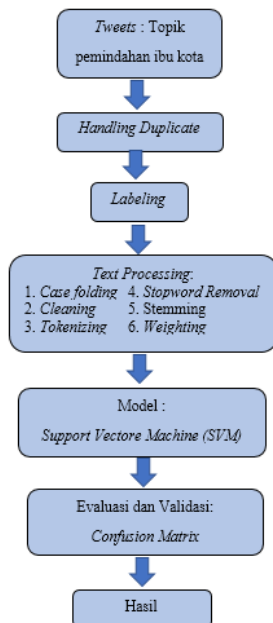
terkait sentimen calon presiden Jokowi dan Prabowo untuk dilakukan analisis sentimen dengan algoritma SVM, hasil akurasi yang diperoleh dari analisis sentimen yang dilakukan adalah 86% (Lukmana, Subanti dan Susanti, 2019). Sementara itu pada tahun 2019, Utama dkk melakukan analisis sentimen terkait dengan efektifitas kebijakan pemerintah dalam Sistem Ganjil Genap di Tol Bekasi menggunakan SVM. Data yang diperoleh berasal dari *multiplatform* media sosial; *twitter*, *youtube*, *facebook* dan *instagram*. Total dataset yang digunakan adalah 440 dataset dengan polaritas 220 positif dan 220 negatif. Penelitian tersebut menghasilkan akurasi sebesar 78,18% (Utama dkk., 2019). Analisis kebijakan pemerintah lainnya juga dilakukan oleh Yanti dkk dimana peneliti mengimplementasikan data *tweet* terkait dengan kebijakan Ganjil Genap Kendaraan bermotor di Ibu Kota Jakarta dengan BM25 dan K-NN. BM25 diimplementasikan pada *output* data hasil *preprocessing*, selanjutnya hasil komputasi BM25 diklasifikasikan kembali menggunakan K-NN. Akurasi yang dihasilkan dengan menggunakan gabungan 2 metode ini adalah 66,5% (Yanti, Indriati dan Adikara, 2019). Penelitian selanjutnya dilakukan oleh Mas'udah dkk pada tahun 2020 menggunakan obyek penelitian topik pemindahan ibu kota. Analisis sentimen dengan topik tersebut mengimplementasikan algoritma *Naive Bayes Classifier* untuk mengklasifikasikan sentimen positif, negatif dan netral. Akurasi yang diperoleh dari klasifikasi menggunakan metode *Naive Bayes Classifier* adalah 64,40% (Mas'udah, Dyar dan Anjani, 2020).

Berdasarkan tinjauan literatur dari penelitian sebelumnya, penulis mengusulkan metode SVM untuk diterapkan pada *tweets* topik pemindahan ibu kota Indonesia untuk tujuan klasifikasi kelas sentimen pada media sosial *twitter*. *Support Vector Machine (SVM)* merupakan metode terbaik diantara beberapa metode lainnya karena mampu mengkomputasi data dengan dimensi tinggi sehingga tingkat akurasi yang dihasilkan lebih baik (Siringoringo dan Jamaluddin, 2019). Teknis klasifikasi dilakukan dengan cara mengklasifikasikan menjadi 2 kelas yakni positif dan negatif. Dengan penerapan SVM ini diharapkan mendapatkan hasil akurasi yang baik serta dapat diketahui kecenderungan sentimen publik yang terdapat pada kontroversi wacana pemindahan ibu kota Indonesia di media sosial *twitter*.

2. METODE PENELITIAN

Pada penelitian ini, penulis menggunakan metode eksperimen/percobaan. Metode eksperimen adalah metode sistematis yang menghubungkan sebab-akibat. Penelitian ini tergolong penelitian kuantitatif yang didalamnya mengandung syarat; *control*, *manipulate* dan *observation* (Nazir, 2017). Eksperimen dilakukan untuk mendapatkan model terbaik dengan mengimplementasikan *tweets* dengan metode SVM guna analisis sentimen terkait topik pemindahan ibu kota. Validasi model dilakukan

dengan menggunakan *k-fold cross validation*. Adapun evaluasi model dilakukan dengan mengukur tingkat akurasi menggunakan *Confusion Matrix*. Gambar 1 berikut ini merupakan metode yang diusulkan dalam penelitian ini.



Gambar 1. Metode Penelitian

2.1. Dataset

Pada penelitian ini digunakan data *tweets* terkait topik pemindahan ibu kota Indonesia yang diambil dengan metode *crawling* pada API *twitter*. Tabel 1 berikut merupakan *sample tweets* yang digunakan.

2.2. Crawling

Crawling merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan *Uniform Resource Locator (URL)*. URL ini menjadi acuan untuk mencari semua *hyperlink* yang ada pada website. Kemudian dilakukan *indexing* untuk mencari kata dalam dokumen pada setiap *link* yang ada. Data yang diambil merupakan *tweets* yang terdapat dalam media sosial twitter dengan pemanfaatan media API twitter. Akses *tweets* membutuhkan hak akses berupa *consumer key*, *consumer secret*, *access token* dan *access token secret*. Kata kunci yang menjadi *keyword* adalah sebagai berikut: pemindahan ibu kota, ibu kota pindah, pemindahan ibu kota jakarta dan ibu kota jakarta. *Crawling* data dilakukan dari tanggal 17 Januari 2020 sampai dengan 20 Februari 2020 dengan total perolehan *tweets* sejumlah 1.320 *tweets*.

2.3. Preprocessing

Tahap *preprocessing* dilakukan guna pengolahan data mentah menjadi koleksi data yang siap digunakan. Secara garis besar terdapat dua tahapan *preprocessing* dalam penelitian ini, yakni *preprocessing* secara umum dan *text processing*. *Preprocessing* secara umum dilakukan dengan tujuan pembersihan data dari jumlah data yang ganda (*handling duplicate*) sementara *text processing*

dilakukan untuk seleksi data yang akan diproses pada dokumen (Mujilawati, 2016). Guna memaksimalkan *output preprocessing* dari penelitian sebelumnya (Azzahra dan Wibowo, 2020) beberapa tahapan *preprocessing* yang dilakukan dalam penelitian ini yaitu:

1. *Handle Duplicates* : tahap *filtering tweets* hasil *crawling* yang memiliki kemungkinan *double* dalam koleksi data.
2. *Labeling* : Proses *labeling* dilakukan oleh dua orang annotator, dimana *output* dari annotator pertama akan di *cross check* oleh annotator kedua.
3. *Case folding* : tahap merubah teks dalam dokumen menjadi bentuk standar yakni *lowercase*.
4. *Cleaning* : proses penghilangan komponen tertentu yang terdapat dalam *tweet* yakni *Uniform Resource Locator (URL)*, *username*, *RT (Retweet)*, karakter *HTML*, dan *hashtag*.
5. *Tokenize* : proses pemecahan teks menjadi kata, dengan batasan tanda baca dan spasi.
6. *Stopword Removal* : dalam tahap ini akan dilakukan penghapusan kata berdasarkan kata yang terdapat dalam *stoplist*.
7. *Stemming* : Proses penguraian kata menjadi bentuk kata dasar.
8. *Weighting* : proses pembobotan kata dengan *TF-IDF*.

Tabel 1. *Tweets* Topik Pemindahan Ibu Kota

	Tweet
1	Pede banget udah mau main pindahkan ibukota sj. Padahal mayoritas rakyat gak mau pindah. Pak @aniesbaswedan mohon jangan mendukung proyek pencitraan ini sebelum ada hasil survey resmi dari lembaga asing yg independen apakah rakyat mmg menginginkan pemindahan ibukota. cc @ILCtv1 https://t.co/kJhaFeiFmw
2	"RT @KangFathanNow: Anggaran Pemindahan Ibukota Negara harus terukur.. #SuaraParlemen https://t.co/xAg6sVNC36 "
3	Presiden Jokowi Gelar Ratas Lanjutan Rencana pemindahan Ibukota https://t.co/Jb0GSYXVOn
4	@asumsico BPIP tu kerjanya ngapain sih.. mending bubarin aja, terus anggarannya dialihkan aja ke pemindahan ibukota.
5	Konsep pemindahan ibukota RI sdh Final, seandainya Presiden @jokowi lengser sekalipun pengantinya harus meneruskan , karena kita bukan ingin membangun ibukota parpol tapi sebuah ibukota Indonesia
.....
1.235	@erwannusantara Kalah pamor dengan Preaiden @aniesbaswedan dalam mengatasi permasalahan Jakarta kok terus pindah Ibukota, oalah jooook...jook...!
1.236	Konsep pemindahan ibukota RI sdh Final, seandainya Presiden @jokowi lengser sekalipun pengantinya harus meneruskan , karena kita bukan ingin membangun ibukota parpol tapi sebuah ibukota indonesia

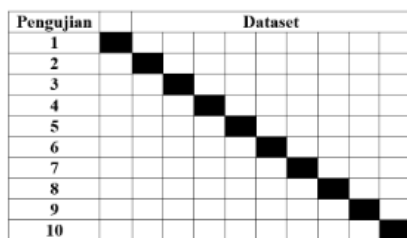
2.4. Rasio Pembagian data Training dan Testing

Pembagian data dilakukan untuk menentukan data *training* dan *testing* dengan perbandingan 9:1 (Han, Kamber dan Pei, 2012). Adapun data *training* yang digunakan adalah 90% dari total koleksi data

yang telah mengalami proses *handling duplicate*, 10% lainnya adalah data *testing*.

2.6. Validasi dan Evaluasi

Validasi dilakukan untuk mengetahui *performace* model yang diusulkan, dalam tahap ini dilakukan validasi model *k-fold Validation* (Alkhaldi dkk., 2020). *K-fold validation* yaitu salah satu model *cross validation* yang bekerja dengan membagi data sejumlah nilai *k* dan iterasi/perulangan sebanyak nilai *k* pula. Gambar 2 berikut merupakan ilustrasi model *10-fold validation*. Eksperimen dengan nilai *k*=10 dilakukan untuk mendapatkan nilai akurasi terbaik. Iterasi berlangsung 10 kali, adapun variasi *training* dan *testing* menggunakan kombinasi 10 bagian. Sementara Evaluasi dilakukan dengan *confusion matrix*.



Gambar 2 Ilustrasi model *10-fold validation*

3. HASIL DAN PEMBAHASAN

Handling duplicate dilakukan guna menghapus sentimen yang kemungkinan ganda. Biasanya *tweets* ganda dihasilkan dari *retweet*. Data *output* dari proses *handling duplicate* berjumlah 1.236 *tweets* dari koleksi data sebelumnya yakni 1.320 *tweets*, data inilah yang nantinya digunakan dalam proses selanjutnya. Tahap *Labeling* data *tweets* dilakukan secara manual oleh dua orang annotator. Annotator 1 bertugas mengklasifikasikan secara manual sentimen positif dan negatif sedangkan annotator 2 bertugas melakukan *cross check* hasil klasifikasi sentimen annotator 1. Penelitian ini terdiri dari 2 polaritas yakni “positif” yang bermakna mendukung wacana pemerintah terkait topik pemindahan ibu kota dan “negatif” yang bermakna menolak atau mengkritisi wacana pemerintah terkait wacana pemindahan ibu kota. Dari proses pelabelan didapatkan data 404 *tweets* positif dan 832 *tweets* negatif (ditunjukkan dalam Gambar 3). Sedangkan *sample* hasil *labeling* yang dilakukan oleh annotator tersaji dalam Tabel 2.

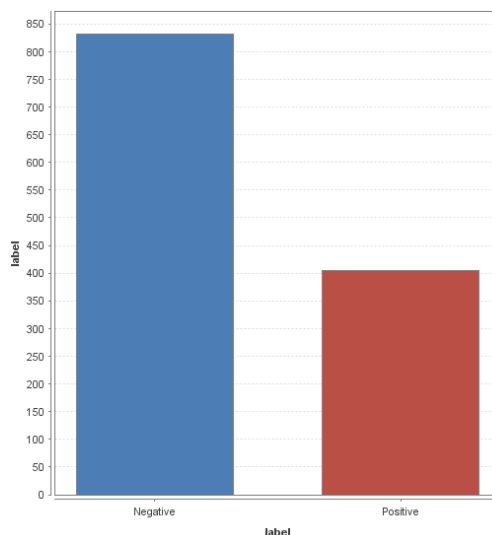
Tabel 2. Hasil *labeling* sentimen positif dan negatif

Tweetsf	Sentimen
@tembangjiwa Itupun jubir yang menegaskan Fajrul Rahman orang kelahiran Banjarmasin. Makanya ibukota dipindah ke Kalimantan mungkin supaya cina bisa jadi presiden.	Negatif
@hakaemel Perubahan iklim bsa di atasi? Gak salah tuh? Hutan kalimantan kebakar diem aja dy, tmbah lg mau pindah	Negatif

Tweetsf	Sentimen
ibukota k kalimantan, apa gak rusak hutan jg tuh? Yaa lagi2 hrs maklum deh	
Ekonomi Indonesia 2019 bertumbuh 5,02% lebih kecil dibanding 2018: 5.17 persen. Jawa sangat dominan, 59,00 persen dari total PDB. Sumatera 21,3%, Kalimantan 8,0%. Distribusi PDB timpang. Itulah Salah Satu alasan CERDAS mengapa IBUKOTA negara perlu pindah.	Positif
Ahok mampu mengurai masalah sedang Anies hanya menambah masalah, untuk saat ini pemindahan ibukota adalah realistis. Jakarta tidak mampu lagi menampung beban 10 tahun kedepan. https://t.co/BIzQbF5nSf	Positif
"@JackVardan Tinggal kita lihat hari2 kedepan... Bakalan imigrasi besar2an ke Indonesia. Semakin terpuruk anak negeri... Ibukota pindah... ASN pindah... Imigran masuk Jakarta... Jadi kota cina..."	Negatif
@ RT @ixan253: Jika dipaksakan pindah ibukota maka @jokowi akan dikenang sbg presiden pembuka pintu habisnya hutan kalimantan. Alih fungsi hutan	Negatif

3.1 Text processing Output

Text processing merupakan tahap terpenting dalam analisis sentimen, proses ini menentukan kualitas data yang menjadi *input* komputasi algoritma SVM. *Output* dari masing-masing proses *text processing* berpengaruh pada proses selanjutnya. Adapun *output text processing* disajikan dalam tabel-tabel dibawah ini.



Gambar 3. Data *tweets* sentimen positif dan negatif setelah *handle duplicates*

3.1 Case Folding

Tahap *case folding* yang dilakukan menghasilkan *tweet* dengan *case standar* yaitu *lower*

case. Tabel 3 berikut merupakan simulasi hasil tahap *case folding*.

Tabel 3 Case folding

Sebelum case folding	Sesudah case folding
@tembangjiwa Itupun jubir yang menegaskan Fajrul Rahman orang kelahiran Banjarmasin. Makanya ibukota dipindah ke Kalimantan mungkin supaya cina bisa jadi presiden.	@tembangjiwa itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden.

3.2 Cleaning

Tahap *cleaning* dilakukan untuk menghilangkan komponen tertentu seperti URL (*Uniform Resource Locator*), *username*, RT (*Retweet*) karakter HTML, dan *hashtag*. Output dari hasil *cleaning* ditunjukkan pada Tabel 4 dibawah ini.

Tabel 4. Tahap Cleaning

Sebelum cleaning	Sesudah cleaning
@tembangjiwa itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden #tolakpindahibukota	itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden

3.4 Tokenizing

Tahap *tokenizing* dilakukan untuk memecah *tweets* menjadi kata dengan batasan tanda baca dan spasi, simulasi hasil *tokenizing* dapat dilihat pada Tabel 5 dibawah ini.

Tabel 5. Tahap Tokenizing

Sebelum tokenizing	Sesudah tokenizing
itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden	itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden

3.5 Stopword Removal

Stopword removal adalah tahap penghilangan kata yang tidak memiliki arti sesuai dengan daftar *stoplist*, dapat dilihat pada Tabel 6.

3.6 Stemming

Tahap penghapusan kata imbuhan awalan dan akhiran. Imbuhan awalan dalam bahasa indonesia adalah me-, mem-, meng-, di- dan lain sebagainya. Imbuhan akhiran dalam bahasa indonesia adalah -i, -nya, -an. Sehingga kata yang dihasilkan adalah kata dasarnya saja, ditunjukkan pada Tabel 7.

Tabel 6. Tahap Stopword Removal

Sebelum Stopword	Sesudah Stopword
itupun jubir yang menegaskan fajrul rahman orang kelahiran banjarmasin. makanya ibukota dipindah ke kalimantan mungkin supaya cina bisa jadi presiden	jubir menegaskan fajrul rahman orang kelahiran banjarmasin. ibukota dipindah kalimantan cina bisa jadi presiden

Tabel 7. Tahap Stopword Removal

Sebelum Stemming	Sesudah Stemming
jubir menegaskan fajrul rahman orang kelahiran banjarmasin. ibukota dipindah kalimantan cina bisa jadi presiden	jubir tegas fajrul rahman orang lahir banjarmasin. ibukota pindah kalimantan cina bisa jadi presiden

3.7 Weighting

Tahap ini merupakan tahap pembobotan kata menggunakan algoritma *Term Frequency – Invers Document Frequency* (TF-IDF). Berikut merupakan simulasia perhitungan bobot dengan algoritma TF-IDF, ditunjukkan pada Tabel 8 dan Tabel 9 berikut.

Tabel 8. Penentuan Idf

token	Kk	Tf			Df	d/df	Idf (log d/df)
		D1	D2	D3			
program	0	1	0	1	2	1,5	0,176
logika	0	1	0	0	1	3	0,477
semantik	1	1	0	1	2	1,5	0,176
ilmu	1	0	1	2	2	1,5	0,176
individu	0	0	1	0	1	3	0,477
transfer	0	0	0	1	1	3	0,477

Tabel 9. Penentuan nilai w (bobot setiap kata)

kk	W		
	D1	D2	D3
0	0,176	0	0,176
0	0,477	0	0
0,176	0,176	0	0,176
0,176	0	0,176	0,352
0	0	0,477	0
0	0	0	0,477

3.8 Eksperimen Model Support Vectore Machine (SVM)

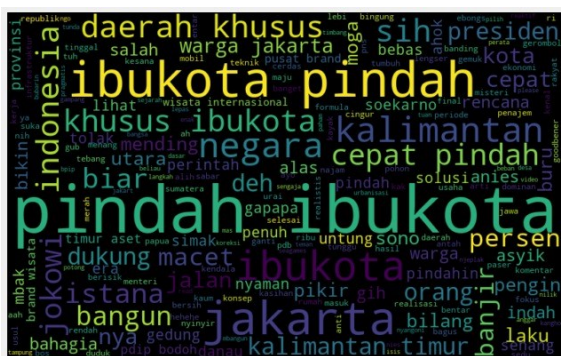
Tujuan penelitian adalah mengimplementasikan algoritma SVM pada sentimen di media sosial twitter terkait topik pemindahan ibu kota untuk mengetahui seberapa besar tingkat akurasi, *recall* dan *precision* dari analisis sentimen yang dilakukan. Komputasi model SVM terhadap bobot tiap kata dalam sentimen sangat menentukan kelas negatif dan positif, serta akurasi yang dihasilkan dari proses pemodelan. Eksperimen dilakukan dengan melakukan *input* kombinasi parameter *C* dan *Convergence Epsilon* dari 0 sampai dengan 1. Tabel 10 berikut merupakan hasil eksperimen kombinasi parameter *Support Vector Machine* (SVM).

Tabel 10. Parameter SVM Training Cycles

C	Epsilon	Accuracy	AUC
0,0	0,0	96.52%	0.986
0,1	0,1	96.68%	0.977
0,2	0,2	96.68%	0.975
0,3	0,3	96.68%	0.976
0,4	0,4	96.68%	0.975
0,5	0,5	96.68%	0.979
0,6	0,6	96.68%	0.972
0,7	0,7	96.68%	0.965
0,8	0,8	96.68%	0.965
0,9	0,9	96.36%	0.960
1,0	1,0	67.31%	0,500
1,0	1,0	67.31%	0,500
1,0	0,0	96.68%	0.972

Berdasarkan Tabel 10 diatas menunjukkan bahwa penerapan metode *Support Vector Machine (SVM)* terbaik adalah dengan kombinasi parameter SVM berupa $C=0,5$ dan $Epsilon=0,5$ dengan hasil akurasi=96,68%.

Guna visualisasi kata yang dominan muncul dalam kelas positif dan negatif berikut disajikan *word cloud* kata yang muncul dalam masing-masing kelas. *Word cloud* kata pada kelas positif ditunjukkan pada Gambar 5, sedangkan *Word Cloud* kelas negatif ditunjukkan pada Gambar 6.



Gambar 5. Word cloud kelas positif



Gambar 6. Word cloud kelas Negatif

3.9 Validasi dan Evaluasi

Sebagaimana disebutkan pada bab sebelumnya, perbandingan data *training* dan *testing* dalam penelitian ini adalah 9:1. Koleksi data yang dimiliki adalah 1.236 *record* sehingga untuk data training berjumlah 1.112 *record* dan *testing* 123 *record*. Pada tahapan validasi dilakukan dengan pengujian *k-fold*

cross validation. Data *training* dibagi acak menjadi beberapa bagian (sesuai nilai *k*) dengan nilai perbandingan sama besar selanjutnya *error rate* dihitung per bagian, rata-rata *error rate* keseluruhan dihasilkan dari penjumlahan *error rate* perbagian. Eksperimen nilai *k* dilakukan sebanyak 9 kali dari nilai $k=2$ sampai dengan $k=10$, seperti yang terlihat pada Tabel 11 dibawah ini.

Tabel 11. Eksperimen parameter SVM dengan validasi *k-fold*

k-fold	Accuracy
2	93.77%
3	96.20%
4	96.44%
5	96.68%
6	96.20%
7	96.52%
8	95.96%
9	96.60%
10	96.12%

Berdasarkan Tabel 11 diatas, penelitian ini menetapkan metode pengujian *5-fold validation*. Proses pengujian diulang sebanyak 5 kali sedangkan hasil pengukuran *error rate* rata-rata dari 5 kali proses pengujian. Berdasarkan eksperimen yang ekstensif serta pembuktian teoritis Tabel 11 diatas, *5-fold* merupakan nilai terbaik guna memperoleh validasi yang akurat.

Pada tahapan Evaluasi hasil, dilakukan evaluasi *output* hasil *text mining*. *Output* tersebut dibandingkan dengan tujuan *business understanding*, dengan demikian dapat diketahui sejauh mana *output text mining* ini dengan capaian tujuan. Hasil evaluasi yang didapatkan melalui model SVM menghasilkan *Confusion Matrix* dengan akurasi=96,68%, *precision*=95.82%, *recall*=94.04% dan $AUC = 0,979$. Seperti yang terlihat pada gambar 7 berikut.

☒ Table View ☐ Plot View

accuracy: 96.68% +/- 0.60% (micro average: 96.68%)

	true Negative	true Positive	class precision
pred. Negative	815	24	97.14%
pred. Positive	17	380	95.72%
class recall	97.96%	94.06%	

Gambar 7. Akurasi Model SVM

Perfomance diukur dengan akurasi dan *Area Under Curve (AUC)* yang ditampilkan dalam bentuk kurva *ROC*. Berikut adalah tampilan *AUC* dari model SVM yang dapat dilihat pada Gambar 8. *Curve Optimis* yang dihasilkan dari model SVM dapat dilihat pada Gambar 9, serta *AUC Pesimistis* model SVM ditunjukkan pada Gambar 10 berikut.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil dari proses pengujian yang telah dilakukan terhadap *tweets* sentimen pemindahan ibu kota dari media sosial twitter sebanyak 1.236 *tweets* (404 positif dan 832 negatif) menggunakan SVM maka hasil hipotesa awal sesuai hasil akhir yakni SVM lebih baik dari metode sebelumnya (BM25 + KNN dan *Naive Bayes*). Dengan akurasi =96,68%, *precision*=95.82%, *recall*=94.04% dan AUC = 0,979.

Pada data sentimen wacana pemerintah terkait topik pemindahan ibu kota dibutuhkan usaha yang besar pada tahap *text processing* awal yakni pada bagian *labeling*. Penanganan sentimen negasi dalam bahasa Indonesia belum menjadi fokus dalam penelitian ini, sehingga sentimen dengan kata negasi belum dapat ditentukan polaritasnya dengan optimal. Sebagai perbaikan dari penelitian selanjutnya perlu dikaji lebih mendalam perbandingan hasil analisis sentimen dengan menerapkan penanganan negasi dengan yang tidak mengalami penanganan negasi dalam bahasa Indonesia, misalnya dengan menggunakan algoritma *Firs Sentiment Word* (FSW), *Fixed Window Length* (FWL) atau *Rest of Sentences* (RoS).

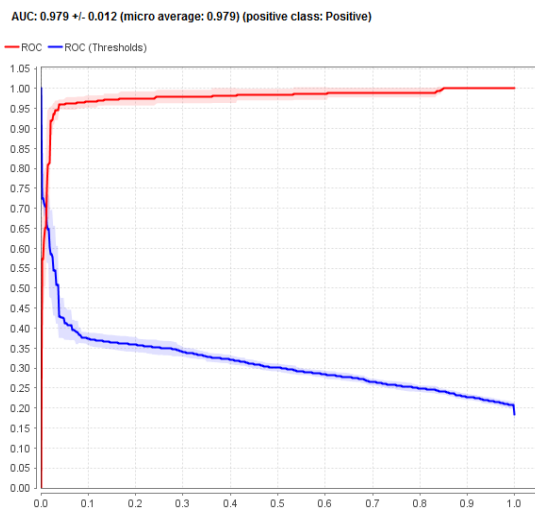
UCAPAN TERIMAKASIH

Ucapan terimakasih yang mendalam penulis sampaikan kepada KEMENRISTEK/BRIN atas pendanaan pada skema Penelitian Dosen Pemula (PDP) tahun 2020.

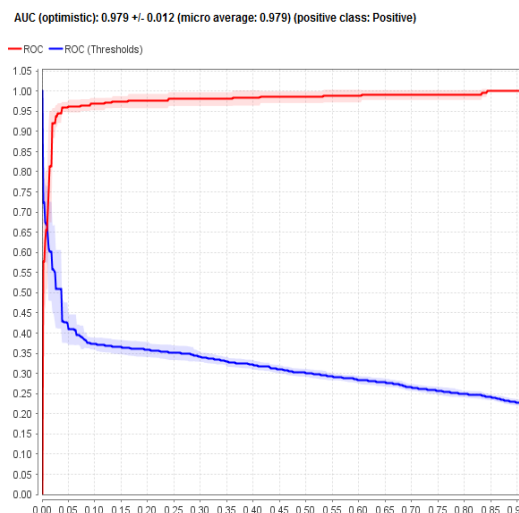
DAFTAR PUSTAKA

- ALKHALDI, S., ALZUAIBI, S., ALQAHTANI, R., ALSHAMMARI, A., ALYOUSIF, F., ALBOANEEN, D.A. & ALMELIHI, M., 2020. Twitter Sentiment Analysis on Activities of Saudi General Entertainment Authority. *Journal of Auckland University of Technology*, pp.2–6.
- ANDRIANI, F., 2018. FENOMENA SOCIAL CIMBER MELALUI TWITWAR. *Jurnal Pustaka Komunikasi*, 1(2), pp.349–360.
- AZZAHRA, S.A. & WIBOWO, A., 2020. Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi Dengan Algoritme Naïve Bayes Untuk Ulasan Wisata Kuliner Pada Multi-Aspect Sentiment Analysis Based On Emoticon Conversion With Naïve Bayes Algorithm For Culinary Tourism Review On Tripadvisor Web. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(4), pp.737–743.
- BRAHIMI, B., TOUAHRIA, M. & TARI, A., 2019. Improving sentiment analysis in Arabic: A combined approach. *Journal of King Saud University - Computer and Information Sciences*. [online] Available at: <<https://doi.org/10.1016/j.jksuci.2019.07.011>>

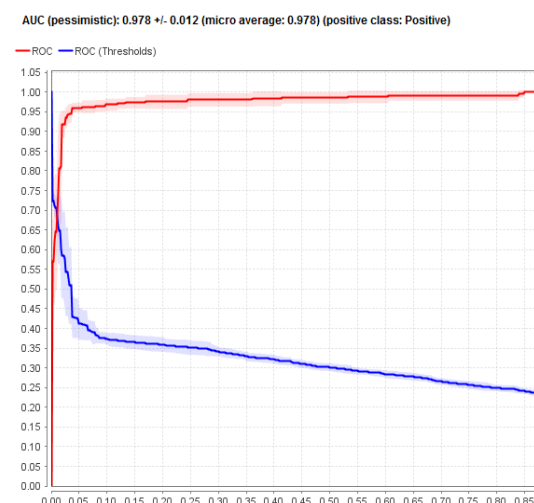
FAUZI, M.A., 2018. Random Forest Approach fo



Gambar 8. AUC model SVM yang dihasilkan



Gambar 9. AUC Optimistic Model SVM



Gambar 10 AUC Pessimistic Model SVM

- Sentiment Analysis in Indonesian Language. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), pp.46–50.
- HAN, J., KAMBER, M. & PEI, J., 2012. *Data mining: concepts and techniques*. Third Edit ed. [online] *Morgan Kaufmann*, USA: Elsevier. Available at: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Data+mining+concepts+and+techniques#1%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Data+mining:+concepts+and+techniques#1%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Sea>>.
- KUSEN, E. & STEMBECK, M., 2018. Politics , sentiments , and misinformation : An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. *Online Social Networks and Media*, [online] 5, pp.37–50. Available at: <<https://doi.org/10.1016/j.osnem.2017.12.002>>.
- LUKMANA, D.T., SUBANTI, S. & SUSANTI, Y., 2019. Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter. In: *Seminar Nasional Penelitian Pendidikan Matematika*. pp.154–160.
- MAS'UDAH, E., DYAR, E. & ANJANI, A., 2020. Analisis Sentimen: Pemindahan Ibu Kota Indonesia Pada Twitter. *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, 1(2), pp.397–401.
- MUJILAHWATI, S., 2016. Pre-Processing Text Mining Pada Data Twitter. Yogyakarta: SENTIKA.pp.49–56.
- NAZIR, M., 2017. *Metode Penelitian*. Bogor: Ghalia Indonesia.
- NOVANTIRANI, A., SABARIAH, M.K. and EFFENDY, V., 2015. Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. In: *e-Proceeding of Engineering*. pp.1177–1183.
- ÖZTURK, N. & AYZAZ, S., 2017. Sentiment Analysis on Twitter : A Text Mining Approach to the Syrian Refugee Crisis. *Telematics and Informatics*, [online] (October). Available at: <<https://doi.org/10.1016/j.tele.2017.10.006>>.
- PANDEY, A.C., RAJPOOT, D.S. & SARASWAT, M., 2017. Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing and Management*, 53, pp.764–779.
- PRATAMA, S.F., ANDREAN, R. & NUGROHO, A., 2019. Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis. *Journal of Information Technology and Computer Science*, 4(2).
- PRIYATNO, A.M., MUTTAQI, M.M., SYUHADA, F. and ARIFIN, A.Z., 2019. Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy dan Global Vectors for Word Representations Tweet ' s Hashtag. *Jurnal Ilmiah Teknologi Sistem Informasi*, 1(Juni), p.37.
- QIU, J., LIN, Z. & SHUAI, Q., 2019. Investigating the Opinions Distribution in the Controversy on Social Media. *Information Sciences*. [online] Available at: <<https://doi.org/10.1016/j.ins.2019.03.041>>.
- ROZI, I.F., PRAMONO, S.H. & DAHLAN, E.A., 2012. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), pp.37–43.
- SEPTIANA, D., 2018. Palangka Raya the Capital City of Indonesia : Critical Discourse Analysis on News about Moving the Capital City from Jakarta. *Advances in Social Science, Education and Humanities*, 280, pp.190–202.
- SIRINGORINGO, R. and JAMALUDDIN, 2019. Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online. *Jurnal Penelitian Teknik Informatika Universitas Prima Indonesia (UNPRI) Medan*, 2(2), pp.1–6.
- TERAN, L. & MANCERA, J., 2019. Dynamic profiles using sentiment analysis and twitter data for voting advice applications. *Government Information Quarterly*, [online] (February 2018), pp.1–16. Available at: <<https://doi.org/10.1016/j.giq.2019.03.003>>.
- THAKUR, R.K. & DESHPANDE, M. V., 2019. Kernel Optimized-Support Vector Machine and Mapreduce framework for sentiment classification of train reviews. *Indian Academy of Science*, [online] 44(1), pp.1–14. Available at: <<https://doi.org/10.1007/s12046-018-0980-1>>.
- TOUN, N.R., 2018. Analisis Kesiapan Pemerintah Provinsi Kalimantan Tengah dalam Wacana Pemindahan Ibu Kota Negara Republik Indonesia ke Kota Palangkaraya. *Jurnal Academia Praja*, 1, pp.129–148.
- Utama, H.S., Rosiyadi, D., Prakoso, B.S. and Ariadarma, D., 2019. Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma Support Vectore Machine. *Jurnal RESTI*, 1(10), pp.2–8.
- WANG, Q., LIU, K. & MA, K., 2019. Emotional Analysis of Public Opinions in Colleges and Universities: Based on Naive Bayesian Classification Method. *Journal of Physics*.
- YANTI, D.S.A., INDRIATI & ADIKARA, P.P., 2019. Analisis Sentimen Tentang Kebijakan Ganjil Genap Kendaraan Bermotor di DKI Jakarta Pada Twitter Menggunakan BM25 dan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK)*, 3(3), pp.2626–2631.
- ZUHDI, A.M., UTAMI, E. & RAHARJO, S., 2019.

- Analisis Sentimen Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-Nn. *Jurnal Informa Politeknik Indonusa Surakarta*, 5(2), pp.1–7.
- ZULFA, I. and WINARKO, E., 2017. Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network. *IJCCS*, 11(2), pp.187–189.

Halaman ini sengaja dikosongkan