Jurnal Serina Sains, Teknik dan Kedokteran
Vol. 1, No. 1, Feb 2023: hlm 11-17 ISSN-L 2986-5204 (Versi Elektronik)

PENDETEKSI UJARAN KEBENCIAN PADA *PLATFORM* MEDIA SOSIAL TWITTER MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Brian Wijaya[1] & Viny Christanti Mawardi[2]

[1]Pogram Studi Sarjana Teknik Informatika, Universitas Tarumanagara Jakarta
Email: Brian.535200069@stu.untar.ac.id
[2]Fakultas Teknologi Informasi, Universitas Tarumanagara Jakarta
Email: vinym@fti.untar.ac.id

ABSTRACT

Twitter is one of the world social media giants, which has enormous flow text-based comment in every second. There are many types of writing sentiments that users create to discuss something such as famous figures, companies or politics. One of the types is hate speech. The analysis was carried out using the Support Vector Machine method as a text analysis model with the help of TF-IDF to assess the weight of each word. The experiment was carried out with several types of kernels and resulted in varying degrees of accuracy. The types of kernels tested were linear, radial basis function, polynomial and sigmoid with a test data distribution of 20%, 25% and 30%.

Keyword: *Nlp, twitter, support vector machine, hate speech.*

ABSTRAK

Twitter merupakan salah satu media social raksasa dunia yang memiliki jumlah arus ulasan dalam bentuk teks yang sangat besar setiap detiknya. Banyak sekali jenis sentimen tulisan yang dibuat pengguna untuk membahas suatu hal seperti tokoh terkenal, perusahaan atau politik. Oleh karena itu Salah satu jenis dari berbagai jenis sentimen tersebut adalah ujaran kebencian. Analisis dilakukan dengan menggunakan metode *Support Vector Machine* sebagai model analisis teks dan dengan bantuan TF-IDF untuk menilai bobot setiap kata. Uji coba dilakukan dengan beberapa jenis kernel dan menghasilkan tingkat akurasi yang beragam. Jenis jenis kernel yang diuji coba adalah linear, radial basis function, polynomial dan sigmoid dengan pembagian data uji 20%, 25% dan 30%.

Kata kunci: Nlp, twitter, support vector machine, ujaran kebencian.

1. PENDAHULUAN

Pada masa globalisasi, perkembangan teknologi informasi meningkat secara eksponensial. Hal itu mempengaruhi bagaimana cara orang-orang bersosialisasi antar satu dengan yang lainnya, salah satu caranya adalah dengan menggunakan media sosial twitter (Buntoro, 2016). Twitter merupakan salah satu situs media sosial dengan arus data tweet lebih dari 400 juta tweet perhari (Badjatiya et al., 2017). Dikarenakan twitter merupakan platform media sosial yang gratis dan bebas dipergunakan, menjadikan media sosial tersebut digunakan dari berbagai golongan usia, ras dan agama. Kebebasan tersebut mengakibatkan tidak terlepasnya banyak pengguna twitter yang secara sengaja mentweet kalimat kalimat yang mengandung ujaran kebencian (Alfina et al., 2017).

Penggunaan ujaran kebencian biasa digunakan sebagai cara untuk menghina atau mengundang orang lain untuk membenci suatu tokoh figur atau suatu organisasi (Pak & Paroubek, 2010). Dampak yang ditimbulkan dari penggunaan kata atau kalimat tersebut yaitu perpecahan, gangguan psikologi, keributan dan lain lainnya.

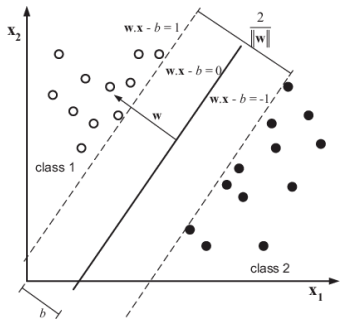
Oleh karena itu pendeteksian ujaran kebencian memiliki peran penting untuk menghindari hal-hal yang mengganggu kehidupan sosial seseorang (Farber, 2012).

Data yang digunakan sebagai penelitian merupakan data tweet dari pengguna twitter berbasis bahasa inggris dengan menggunakan social network scraper sebagai alat untuk pengambilan tweet pengguna. Kemudian data dianalisis menggunakan *Support Vector Machine* (SVM) sebagai metode pembuatan model, yang kemudian diuji pada kernel-kernel seperti polinomial, *radial basis function* (RBF) dan sigmoid untuk membuktikan kernel jenis apa yang memiliki model dengan nilai ketepatan terbaik.

2. METODE PENELITIAN

Support Vector Machine merupakan salah satu metode *supervised learning* yang digunakan untuk melakukan pengklasifikasi data. SVM merupakan metode terbaik untuk melakukan perhitungan data berdimensi tinggi (Tripathy, 2015). dengan konsep dasarnya yang menggunakan *hyperplane* untuk membagi data menjadi dua set. Pada SVM perlu digunakannya batas pemisah yang optimal. Pemisahan batas yang dibutuhkan merupakan batasan yang memiliki jarak maksimal dengan data A dan data B (Pupale, 2018). Hal ini dibutuhkan untuk mendapatkan klasifikasi dengan ketepatan tinggi. Jika garis mendekati salah satu dari data A atau B, batasan tersebut akan memiliki kesalahan saat melakukan klasifikasi. Untuk mendapatkan pemisah tersebut dibutuhkan sebuah alat bantu hitung yaitu Margin maksimum (Kowalczyk, 2017).

Gambar 1
Hyperplane dua dimensi



Rumus pada *hard*-margin SVM terbentuk sebagai berikut:

$$manimum : V(\vec{w}\rightarrow, b) = \frac{1}{2}\vec{w}\rightleftharpoons . \vec{w}\rightarrow, \quad (1)$$

$$subject\ to : \forall_{i=1}^{n} : y_i[\vec{w}\rightarrow . x\rightarrow_i + b] \geq 1, \quad (2)$$

$$\delta = \frac{1}{\| \vec{w}\rightarrow \|} \quad (3)$$

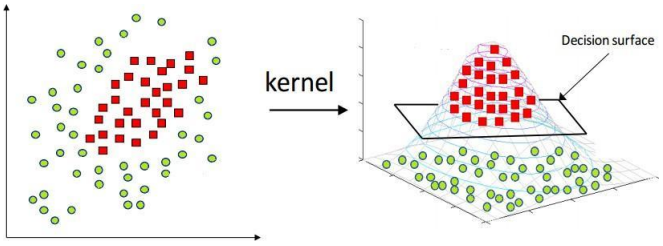
V = *Hyperplane*
$\vec{w}\rightarrow, b$ = Titik yang ditentukan
$\forall^n$ = data titik i
y_i = nilai i di antara 1 dan -1
$\vec{x}\rightarrow_i$ = Jarak titik x_i
δ = Jarak *Hyperplane* dengan data sampel

Menurut Hofmann (2006) tidak semua data dapat dipisahkan secara linear, *Support Vector Machine* menggunakan trik *kernel* untuk memetakan pelatihan vektor ke peringkat ruang dimensi yang lebih tinggi, dengan demikian *hyperplane* dapat ditentukan dan menghasilkan

 https://doi.org/10.24912/jsstk.v1i1.22746

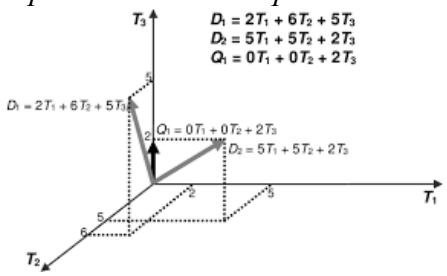
pemisah data yang sempurna seperti pada Gambar 2. Beberapa contoh trik kernel *support vector machine* adalah kernel polynomial, kernel rbf dan kernel sigmoid.

Gambar 2
Trik kernel SVM



Vector Space Model (VSM) merupakan sebuah teknik vektor untuk mencari kemiripan dokumen dengan kata kunci yang ingin dicari. Dengan bantuan t*erm frequency – invese document frequency*, hasil angka tersebut dapat digunakan untuk mendapatkan posisi dokumen pada bidang vektor (Sarkar, 2016). Setelah berhasil mendapatkan bobot dari dokumen dan kata kunci, kedua bobot tersebut akan dihitung menggunakan dot product, dan hasilnya akan dibuatkan peringkat kemiripan. Representasi nilai kemiripan tertinggi itu menunjukan bahawa dokumen merupakan dokumen yang mengandung kata kunci terbanyak.

Gambar 3
Representasi Vector Space Model



Sebuah kumpulan dokumen dapat ditampilkan dalam VSM dalam bentuk matrix yang di mana baris direpresentasikan dengan dokumen-dokumen dan kolom merepresentasikan kata (Usharani & Iyakutti, 2013). Isi dari matrix tersebut berisikan masa bobot(*term weights*) kata yang ada pada setiap dokumen.

Gambar 4
Bentuk Matrix VSM

	T_1	T_2	T_3	$T_{...}$	T_t
D_1	W_{11}	W_{21}	W_{31}	$\square_{...}$	T_{t1}
D_2	W_{12}	W_{22}	W_{32}	$\square_{...}$	T_{t2}
D_3	W_{13}	W_{23}	W_{33}	$\square_{...}$	T_{t3}
$D_{...}$	$\square_{...}$	$\square_{...}$	$\square_{...}$	$\square_{...}$	$\square_{\square}$
D_n	W_{1n}	W_{2n}	W_{3n}	$\square_{...}$	T_{tn}

D pada gambar mengarah pada dokumen teks ke n dan T mengarah pada kata pada semua dokumen dan W merupakan berat kata tersebut pada dokumen teks. Langkah pertama dokumen dihitung terlebih dahulu bobot frekuensi(*term frequency*/TF) yang mengindikasikan banyaknya kemunculan kata pada dokumen menggunakan rumus sebagai berikut:

$$tf_{ij} = \frac{f_{ij}}{max_i\{f_{ij}\}} \quad (3)$$

tf_{ij} = bobot frekuensi kata i pada dokumen j

f_{ij} = frekuensi dari kata i pada dokumen j

max_i = maksimum kata i

Kemudian dilanjutkan dengan menghitung *inverse document frequency*(IDF) dengan rumus di bawah ini:

$$idf_i = \log_2\left(\frac{N}{df_i}\right) \quad (4)$$

idf_i = *inverse document frequency* yang terdapat di i

N = Total banyaknya dokumen

df_i = banyaknya dokumen yang mengandung i

Setelah mendapatkan nilai TF dan nilai IDF akan digambungkan kedua nilai tersebut untuk mendapatkan nilai masa bobot (*term weights*) dengan rumus:

$$Wij = tfij \,.\, idfi \quad (5)$$

Wij = masa bobot kata i pada dokumen j

$tfij$ = bobot frekuensi kata i pada dokumen j

$idfi$ = *inverse document frequency* yang terdapat di i

Setelah mendapatkan nilai masa bobot baru hasil dapat digunakan untuk mencari hasil kemiripan kata dari dokumen dengan *query* yang akan dihitung sebagai *dot product* (Muhajir, 2012):

$$cos(d_j, q) = \frac{dj \cdot q}{\|d_j\|\|q\|} = \frac{\sum WijWiq}{\sqrt{\sum W_{ij}^2}\sqrt{\sum W_{iq}^2}} \quad (6)$$

$sim(d_j, q)$ = kemiripan(dokumen j, *query*)

 https://doi.org/10.24912/jsstk.v1i1.22746

dj = dokumen j

q = *query*

Wij = masa bobot i pada dokumen j

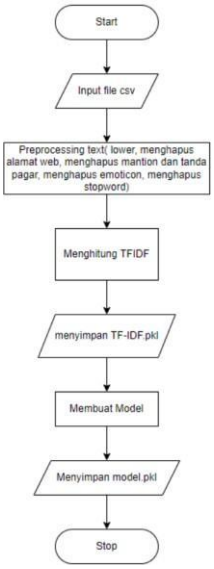
Wiq = masa bobot i pada *query*

3. HASIL DAN PEMBAHASAN

Dibawah ini adalah skema pembuatan model klasifikasi dengan menggunakan data latih

Gambar 5

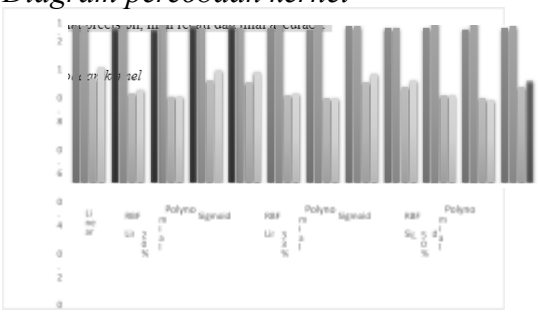
Skema Penelitian



Penelitian dilakukan dengan data yang telah dibagi dengan dua buah jenis label dengan nilai 0 dan 1, yang dimana nilai 1 menunjukan bahwa data tulisan masuk kedalam kategori ujaran kebencian yang bermakna seksisme atau rasisme. Total data latih yang dipakai sebanyak 31962 data dengan label 1 sebanyak 2242 dan label 0 sebanyak 29720. Langkah pertama yang dilakukan dalam keberlaksanaan eksperimen ini adalah dengan melakukan pembersihan teks. Teks yang telah dibersihkan kemudian dihitung matriks bobot TF-IDF-nya dengan menggunakan library sklearn dan disimpan. Pembuatan model dilakukan dengan beberapa metode kernel dan pembagian data latih dan uji 80%:20%, 66.7%:33.3%, dan 50%:50%.

Tujuan eksperimen ini adalah untuk mengetahui kernel apa yang baik digunakan untuk klasifikasi data teks twitter. Hasil pengujian didapat dengan menggunakan confusion matrix untuk menghitung nilai precision, nilai recall dan nilai accuracy.

Gambar 6
Diagram percobaan kernel



Accracy	0.95	0.94	0.93	0.95	0.95	0.94	0.93	0.95	0.94	0.94	0.93	0.94
Precision	0.95	0.97	0.97	0.94	0.94	0.97	0.97	0.95	0.94	0.97	0.97	0.95
Recall	0.63	0.54	0.52	0.62	0.61	0.53	0.51	0.61	0.58	0.53	0.51	0.58
F1-Score	0.7	0.56	0.52	0.68	0.67	0.54	0.51	0.66	0.62	0.53	0.5	0.62

Dari hasil yang telah diuji coba pembagian data terbaik dan kernel yang cocok pada data ini adalah kernel Linear dengan pembagian data latih sebesar 80%:20%, dikarenakan memiliki nilai presisi, recall dan nilai f1 terbesar dan akurasi tertinggi sebesar 95%.

Setelah mendapatkan model klasifikasi terbaik. Kemudian dengan menggunakan social network scraper untuk mendapatkan data tweet yang akan diuji hasil prediksi tweet tersebut dengan model klasifikasi yang telah ditentukan jika hasil prediksi yang didapat menghasilkan nilai 1 maka tweet tersebut mengandung makna seksisme atau rasisme. Dibawah ini adalah percobaan yang dilakukan dengan menggunakan model klasifikasi support vector machine dengan kernel linear dan pembagian data train sebesar 80%:20%.

Gambar 7
Hasil percobaan prediksi model

	Tweet	User	Prediction
0	@SoulDancingStud I will create a Christmas dance	Dancer_nana1	Sexist/Racist
1	@NEWSMAX @Blaidd_tx I am so proud of luz Chene	Christo86017632	Neutral
2	Seth Meyers hit Donald Trump, Eric Trump and Dona	Megresistor	Sexist/Racist
3	What do you call a guy with no arms and no legs who	BotfulJokes	Neutral
4	@Mxstr76ArmyMom I mean PJW's horrible food hor	daleksoup1	Neutral
5	Aubrey O'Day Declares Ex Donald Trump Jr. Has Turn	hardknoxfirst	Neutral
6	@MSNBC Father's will get equal rights in family cour	ImJustHereLook4	Neutral
7	Disrespect Donald Trump	Fuckdonaldxmike	Neutral
8	I just signed a petition urging @meta not to allow	Nocturne_dragon	Neutral
9	Alex jones and Donald Trump âŒ dictators who push	JoshuaKitson4	Neutral

4. KESIMPULAN DAN SARAN

Dari hasil uji coba ini, dapat disimpulkan terdapat beberapa jenis kernel pada metode klasifikasi Support Vector Machine untuk dapat menemukan model terbaik. Perlunya melakukan banyak percobaan pada data seperti kernel dan ratio pemisahan data latih untuk mendapatkan model klasifikasi terbaik serta tingkat akurasi tertinggi. Dan pada pengujian ini hasil model klasifikasi berhasil digunakan dan dapat memberikan hasil prediksi. Saran yang dapat dibagikan penulis adalah pengembangan penelitian selanjutnya menggunakan data dalam bahasa Indonesia dan dapat menambahkan lagi jenis ujaran kebencian. Kemudian dapat mengintegrasikan aplikasi untuk melakukan analisis seberapa buruk tweet pengguna dengan kata kunci tertentu.

https://doi.org/10.24912/jsstk.v1i1.22746

Ucapan Terima Kasih (*Acknowledgement*)
Terima kasih kepada pihak-pihak yang terlibat dalam proses pembuatan artikel ini.

REFERENSI

Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017, October). Hate speech detection in the Indonesian language: A dataset and preliminary study. In *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 233-238). IEEE.

Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017, April). Deep learning for hate speech detection in tweets. In *Proceedings of the 26th international conference on World Wide Web companion* (pp. 759-760), https://doi.org/10.1145/3041021.3054223.

Buntoro, G. A. (2016). Analisis sentimen hatespeech pada twitter dengan metode naïve bayes classifier dan support vector machine. *Jurnal Dinamika Informatika*, *5*(2), 1-12.

Farber, D. (2012, Juni 6). Twitter hits 400 million tweets per day, mostly mobile. *CNET*. https://www.cnet.com/tech/services-and-software/twitter-hits-400-million-tweets-per-day-mostly-mobile/.

Hofmann, M. (2006). Support vector machines-kernels and the kernel trick. *Notes*, *26*(3), 1-16.

Kowalczyk, A. (2017). *Support vector machine succinctly.* Syncfusion.

Muhajir, R. B. (2012). *Metode similarity-mashup untuk framework modul relevant content pada content management system (cms).* Universitas Negeri Sebelas Maret.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).

Pupale, R. (2018). *Support vector machine (svm)-an overview.* Towards Data Science. https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989.

Sarkar, D. (2016). *Text analytics with python: A practical real-world approach to gaining actionable insights from your data.* Apress.

Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of sentimental reviews using machine learning techniques. *Procedia Computer Science*, *57*, 821-829.

Usharani, J., & Iyakutti, K. (2013). A genetic algorithm based on cosine similarity for relevant document retrieval. *International Journal of Engineering Research & Technology (IJERT)*, *2*(2), 1-13.