DETEKSI UJARAN ANCAMAN BERBASIS WEBSITE PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

WEBSITE BASED DETECTION OF THREATS IN SOCIAL MEDIA TWITTER USING SUPPORT VECTOR MACHINE METHOD

Ananda Adhari[1], Muhammad Nasrun S.Si, M.T.[2], Ratna Astuti Nugrahaeni S.T, M.T.[3]

Prodi S1 Teknik Komputer,Fakultas Teknik Elektro, Universitas Telkom
nandaadhari@telkomuniversity.ac.id, muhammadnasrun@telkomuniversity.ac.id, ratnaan@telkomuniversity.ac.id

Abstrak

Ancaman dapat datang darimana saja bahkan dari media sosial, dan termasuk dalam ujaran kebencian. Karena adanya ancaman, muncul kegelisahan dan ketakutan yang pada akhirnya akan meningkatkan kewaspadaan kita pada suatu ancaman. Suatu ancaman dapat datang dalam bentuk apa saja, baik itu ancaman penculikan, ancaman kekerasan, hingga ancaman pembunuhan. Pada umumnya, seseorang yang melakukan tindakan ancaman, identitasnya tidak diketahui atau bersifat *anonymous* dan misterius. Dalam UU ITE terdapat aturan dalam bermedia sosial, yang membuat pengguna media sosial tidak dapat seenaknya melakukan unggahan yang berisi ancaman atau intimidasi hingga persekusi. Pada Tugas Akhir ini, penulis merancang aplikasi website yang digunakan untuk mendeteksi ujaran ancaman pada postingan dari media sosial Twitter. Pada perancangan ini, penulis menggunakan metode dari *machine learning* yaitu *Support Vector Machine*. Dan hasil program pada sistem pendeteksi ujaran ancaman pada postingan Twitter yang dibuat mendapatkan akurasi sebesar 73%, *precision* 72%, *recall* 62,67%, *f-1 score* 61,16%.

Kata kunci : ancaman, media sosial, *machine learning, Support Vector Machine*

Abstract

Threats can come from anywhere even from the social media, and it is also part of hate speech. Because of threat, anxieties and fears arise which ultimately will increase our awareness of a threat. A threat can come in any form, whether it is a threat of kidnapping, threat of violence, or threat of murder. In general, someone who do an act of threat, his identity will remained anonymous and mysterious. In the ITE Law, there are rules in the social media. So that the users of social media cannot uploads that contains a threat carelessly, or intimidation even persecution. In this Final Project, the author designed a website application to detect a threat utterances on Twitter social media. In this design, the authors use a method from machine learning, named Support Vector Machine. And the results of the program on the threat speech detection system on Twitter posts maded to get an accuracy of73 %, precision of 72%, recall of 62,67%, f-1 score of 61,16%.

Keywords: ***threat, social media, machine learning, Support Vector Machine***

1. Pendahuluan [10 pts/Bold]

Saat ini media sosial adalah salah satu tempat dimana orang-orang dapat mengutarakan apapun yang mereka pikirkan. Khususnya di Twitter, yaitu situs media sosial yang berkembang pesat karena para pengguna dapat berinteraksi dengan pengguna lainnya dari komputer atau perangkat *mobile* dimanapun dan kapanpun. Pada tahun 2010 pengguna Twitter yang terdaftar sekitar 160 juta pengguna (Chiang, 2011). Dalam kebebasan berpendapat, media sosial seharusnya menjadi wadah kebebasan bagi masyarakat. Namun, pada kenyataannya pasal – pasal defamasi di Indonesia bisa menjadi halangan bagi iklim demokrasi Indonesia. Seperti pada pasal 27 ayat (3) Undang-undang Informasi dan Transaksi Elektronik dianggap sebagai pasal karet, karena sering dijadikan dasar

hukum untuk menjerat pihak-pihak yang telah dianggap melakukan pencemaran nama baik di media sosial. Dari pencemaran nama baik tak luput dengan hadirnya suatu unsur ujaran ancaman. Ancaman dapat hadir dalam bentuk penculikan, kekerasan, hingga pembunuhan. Hal tersebut sangat mencemaskan masyarakat Indonesia khususnya bagi pengguna media sosial sekarang. Pada Tugas Akhir ini penulis membuat aplikasi website yang berfokus pada perancangan untuk mendeteksi ujaran ancaman pada postingan di media sosial khususnya di Twitter. Penulis akan menggunakan metode dari *machine learning* yaitu metode *Support Vector Machine*. *Support Vector Machine* dipilih karena memberikan performansi yang sangat baik dalam kasus klasifikasi teks yang datanya cukup banyak. Dengan dibuatnya sistem deteksi ini, penulis berharap ujaran ancaman di media sosial Twitter dapat di deteksi dan membantu lembaga terkait untuk menangani kasus-kasus mengenai ancaman yang ada di media sosial Twitter.

2. Dasar Teori /Material dan Metodologi/perancangan

2.1 Ancaman

Ancaman merupakan salah satu hal yang menakutkan dan dihindari oleh orang-orang. Karena dapat membuat efek yang beragam pada korbannya, mulai dari individu atau perorangan hingga bangsa dan negara pun dapat terkena dampaknya. Jika diluruskan, ancaman dapat diartikan sebagai setiap usaha dan kegiatan untuk membahayakan seseorang, dari kelompok hingga bangsa dan negara (Pelayan Publik, 2019).

2.2 Support Vector Machine

Salah satu teknik yang biasa digunakan untuk prediksi dalam mengklasifikasi dan regresi adalah *Support Vector Machine* atau disingkat menjadi SVM. Dalam pengenalan teks, pengenalan objek, dan pengklasifikasian teks, metode SVM sangat sering dipakai [2]. Pada prinsipnya, SVM bersifat linier *classifier* yang dimana artinya klasifikasi yang secara linier dapat dipisahkan. Namun, apabila dimasukkan konsep kernel di ruang kerja berdimensi tinggi, SVM bisa bekerja pada klasifikasi non-linier. Untuk di ruang berdimensi tinggi, nanti akan mencari *hyperplane* yang dapat memaksimalkan jarak antar kelas data.

2.3 Twitter

Twitter adalah media sosial yang berukuran kecil atau bisa disebut juga dengan *micro-blogging* yang dibuat oleh Jack Dorsey pada Maret 2016 dan dirilis pada bulan Juli tahun 2006. Ciri khas dari twitter adalah fitur untuk posting atau men-*tweet* sesuatu dengan ukuran maksimal sebanyak 140 karakter[3].

2.4 Text Pre-Processing

Text preprocessing merupakan proses mengubah bentuk data yang belum memliki struktur menjadi data yang terstruktur sesuai dengan kebutuhan, untuk proses *mining* yang lebih lanjut. Sebuah teks yang ada harus dipisahkan, hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda. Pengubahan bentuk dapat berupa memecah paragraf menjadi kalimat dan kalimat akhirnya menjadi kata serta dapat menghilangkan angka, simbol atau karakter-karater lainya. Tahapan *preprocessing* berdasarkan meliputi: *case folding*, *tokenizing/parsing*, *filtering*, *stemming*[7].

2.5 Klasifikasi Pengujian

Pengujian dilakukan dengan mengukur performansi dari model klasifikasi, diukur dari perbandingan antara data latih dengan dengan data uji. Terdapat 4 parameter untuk melakukan uji performansi, yaitu *accuracy*, *precision*, *recall*¸*dan f-1 score*. Parameter ini didapatkan dengan membandingkan data uji dengan data latih dari hasil validasi dengan Balai Bahasa.

Accuracy merupakan tingkat kedekatan dari data latih dan data uji. Persamaan untuk menghitung *accuracy* yaitu:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.1)$$

Precision merupakan tingkat ketepatan dari data yang diminta oleh pengguna dengan data yang dihasilkan oleh sistem. Persamaannya yaitu:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

Recall merupakan tingkat keberhasilan sistem dalam mengklasifikasikan. Persamaannya yaitu:

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

F1-Score merupakan evaluasi yang terdiri dari gabungan antar *precision* dan *recall.* Persamaannya yaitu:

$$F1\text{- }score = 2\ x\ \frac{precision\ x\ recall}{precision+recall} \quad (2.4)$$

Ket:

TP = Data yang diklasifikasikan sebagai ujaran kebencian

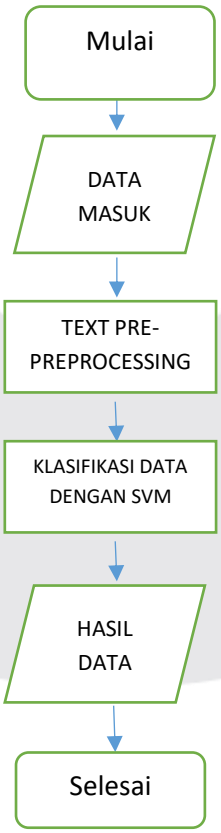
TN = Data yang bukan ujaran kebencian tetapi sistem mengklasifikasikan sebagai ujaran kebencian

FP = Data yang merupakan ujaran kebencian tetapi sistem mengklasifikasikan bukan ujaran kebencian

FN = Data yang diklasifikasikan bukan ujaran kebencian

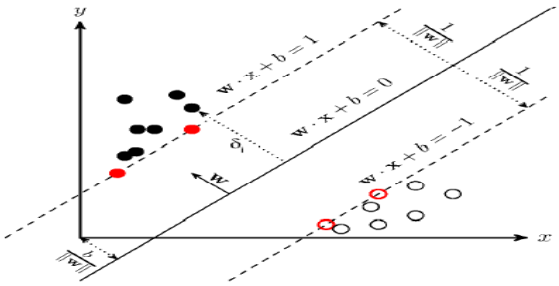
3. Pembahasan

3.1. Perancangan Sistem



Gambar 1. Rancangan Sistem

Langkah pertama yang dilakukan pada sistem sesuai gambar 3.3 adalah memasukkan kata pada *website*, setelah itu di dalam *website* akan melakukan *crawling data* dari Twitter dan kata – kata tersebut akan di proses dengan *Text Preprocessing*. Selanjutnya, akan di klasifikasi menggunakan metode SVM dan nanti akan muncul hasil postingan dari Twitter yang berkaitan dengan kata – kata yang telah dimasukkan pada *website* sebelumnya.



3.2 Konsep Dasar Support Vector Machine

Metode SVM dapat dipertimbangkan dalam pengaturan sebuah klasifikasi biner. Jika ada contoh data *training* atau data latih {**x**1 . . . **x**n} yang merupakan vector di beberapa ruang $X \subseteq \mathrm{R}d$. Dan juga ada contoh {$y1$ . . . yn } yang dimana $yi \in \{-1, 1\}$.

Gambar 2. *Hyperplane* membagi sebuah kelas

Dalam bentuk sederhana nya, SVM adalah *hyperplane* yang memisahkan data *training* dengan margin yang maksimal. Dan data training yang mempunyai jarak terdekat dengan *hyperplane* disebut dengan *support vectors*.

Persamaan pada *Hyperplane* SVM dinyatakan sebagai berikut:

$$w.x + b = 0 \qquad (3.1)$$

Dimana *w* adalah *normal* dari *hyperplane* dan $\frac{b}{\|w\|}$ adalah jarak *hyperplane* ke titik *origin*[14]. Pada titik yang berada di *class x* adalah titik data yang memenuhi persamaan

$$[(x_i, w) + b] \geq -1 \; untuk \; y_i = -1 \qquad (3.2)$$

Lalu untuk titik yang berada di *class y* adalah titik data yang memiliki persamaan

$$[(x_i, w) + b] \geq 1 \; untuk \; y_i = +1 \qquad (3.3)$$

4. Pengujian dan Implementasi

Tabel 4.1 Rangkuman Pengujian Partisi Data

Pengujian ke-	Data uji (%)	Data latih (%)	*Precision* (%)	*Recall* (%)	*F-1 Score* (%)	*Accuracy* (%)
1	10	90	72	66	66	71
2	20	80	71	64	64	71
3	40	60	73	67	68	73
4	60	40	72	66	66	71
5	80	20	72	59	56	67
6	90	10	72	54	47	64

Pada tabel 4.1 diatas dapat disimpulkan bahwa pada pengujian pertama hingga ketiga terdapat peningkatan pada *accuracy* yang awalnya dari 71% meningkat hingga 73%. Namun, saat pengujian keempat dan seterusnya mengalami penurunan hingga 64%.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil pengujian yang telah dilakukan dapat disimpulkan bahwa:

1. Performansi pengujian akurasi terbaik yang didapatkan sebesar 73%. Banyaknya dataset dan proses dalam *pre-processing* dapat mempengaruhi hasil yang di dapat.
2. Sistem deteksi ujaran ancaman dalam Bahasa Indonesia pada postingan tweet di Twitter dengan metode *Support Vector Machine* berhasil untuk melakukan klasifikasi kalimat yang mengandung ancaman dan bukan ancaman.

Pada proses pengujian sistem diperoleh nilai rata-rata untuk parameter *precision*, *recall*, dan *f-1 score* dengan nilai masing – masing 72%, 62,67%, dan 61,16%.

5.2 Saran

Saran yang dapat diusulkan untuk penelitian lebih lanjut kedepannya adalah:

1. Dataset yang digunakan harus balance dalam implementasinya.

Proses dalam *pre-processing* yang digunakan serta penggunaan klasifikasi harus sesuai dengan data agar hasil yang di dapat lebih memuaskan.

Reference:

1. Ma'rifah, Hidayatul & Wibawa, Aji & Akbar, Muhammad. (2020). **Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing**. Sains, Aplikasi, Komputasi dan Teknologi Informasi. 2. 70. 10.30872/jsakti.v2i2.2681**.**
2. Simon Tong, 2001. **Support Vector Machine Active Learning with Aplications to Text Classification.** Stamford: *Journal of Machine Learning Research Computer Science Department Stamford University.*
3. Teguh Wahyono, 2018. ***Fundamental Of Python For Machine Learning.*** Salatiga: Penerbit Gava Media.
4. Sembodo, J. Eka, E. Budi Setiawan, and ZK Abdurahman Baizal, 2016. **Data Crawling Otomatis pada Twitter**. *Indonesian Symposium on Computing (Indo-SC).*

5. Byun, Changhyun, Hyeoncheol Lee, and Yanggon Kim, 2012. **Automated Twitter data collecting tool for data mining in social network**. *Proceedings of the 2012 ACM Research in Applied Computation Symposium*.
6. Zhang, Wen, Taketoshi Yoshida, and Xijin Tang, 2008. **Text Classification Based on Multi-word with Support Vector Machine**. *Knowledge-Based Systems* 21.8: 879-886.
7. Erizal, Elvira, Budhi Irawan, Casi Setianingsih, 2019. **Hate Speech Detection in Indonesian Language on Instagram Comment Section Using Maximum Entropy Classification Method.** International Conference on Information and Communications Technology (ICOIACT).
8. Colas, Fabrice, and Pavel Brazdil, 2006. **Comparison of SVM and some older classification algorithms in text classification tasks**. *IFIP International Conference on Artificial Intelligence in Theory and Practice*. Springer, Boston, MA.
9. Qiu, Lin & Lin, Han & Ramsay, Jonathan & Yang, Fang. (2012). **You are what you tweet: Personality expression and perception on Twitter**. *Journal of Research in Personality*. 46. 710–718.
10. Franky, & Manurung, R. (2008). **Machine Learning-based Sentiment Analysis of Automatic Indonesian Translations of English Movie Reviews**.
11. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, 2002. **Thumbs up? Sentiment Classification using Machine Learning Techniques**. *Proceedings of EMNLP 2002*
12. Buntoro, 2017. **Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter.** *Integer Journal*, Vol 2, No 1, Maret 2017.
13. Purnamawan, I**.** (2015)**. SUPPORT VECTOR MACHINE PADA INFORMATION RETRIEVAL.** Jurnal Pendidikan Teknologi dan Kejuruan**.** 12. 10.23887/jptk-undiksha.v12i2.6481**.**
14. Catur, H. (2018**). Perbandingan Metode Support Vector Machine (SVM) Linear, Radial Basis Function (RBF), dan Polinomial Kernel Dalam Klasifikasi Bidang Studi Lanjut Pilihan Alumni UII.**
15. Marlin, D. **Analisis Sentimen Pada Komentar Akun Perfilman Instagram Dengan Metode Maximum Entropy**. pp. 5, 2020.
16. Badan Pengembangan dan Pembinaan Bahasa. 2019. Kamus Besar Bahasa Indonesia Daring.
17. Prawiro, M. 2019. **Pengertian Ancaman: Arti, Jenis-Jenis, dan Contoh Ancaman**.
18. Pelayanan Publik. 2019. **Ancaman, Jenis dan Dampaknya Bagi Individu Serta Negara**.