BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

Implementasi Algoritma Metode Naive Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter

Ahmad Turmudi Zy, Wahyu Hadikristanto
Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Kabupaten Bekasi, Indonesia
Email: [1]turmudi@pelitabangsa.ac.id, [2]wahyu.hadikristanto@pelitabangsa.ac.id
Email Penulis Korespondensi: turmudi@pelitabangsa.ac.id

Abstrak− *Data breaches dan data leakage* merupakan masalah yang semakin meningkat dalam dunia teknologi saat ini. Ada banyak diskusi yang terjadi di media sosial, terutama di Twitter, mengenai masalah ini. Dalam paper ini, kami menggunakan analisis sentimen untuk mengevaluasi persepsi publik terhadap masalah keamanan data dan kebocoran data di Twitter. Kami menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes untuk melakukan klasifikasi sentimen pada teks yang diambil dari Twitter. Hasil dari penelitian ini akan memberikan gambaran yang lebih baik mengenai persepsi publik terhadap masalah keamanan data dan kebocoran data dan dapat digunakan oleh pihak yang berkepentingan untuk meningkatkan keamanan data Selanjutnya, paper ini akan mendeskripsikan metodologi yang digunakan dalam pengumpulan data dari Twitter, proses pre-processing data, dan implementasi algoritma SVM serta Naïve Bayes untuk melakukan klasifikasi sentimen. Kemudian, hasil dari analisis sentimen akan dianalisis dan dibahas secara mendalam, serta dibandingkan dengan hasil analisis sentimen yang dilakukan dengan metode lain. Pada paper ini didapatkan hasil perhitungan algorithma NaïveBayes dengan *precision* 97%, dan perhitungan dari algorithma SVM didapatkan hasil *Precision* sebesar 80%.Selain itu, paper ini juga akan mencakup kesimpulan dan saran untuk penelitian selanjutnya dalam bidang yang sama.

Kata Kunci: Data breaches; Data leakage; Teknologi; Media Sosial; Twitter; Analisis sentimen

Abstract− Data breaches and data leakage are increasingly becoming a problem in the technology world today. There are many discussions happening on social media, particularly on Twitter, about this issue. In this paper, we use sentiment analysis to evaluate the public perception of data security and data breaches on Twitter. We use the Support Vector Machine (SVM) and Naïve Bayes algorithm to classify sentiment in text taken from Twitter. The results of this research will provide a better understanding of the public perception of data security and data breaches and can be used by stakeholders to improve data security. Additionally, this paper will describe the methodology used in collecting data from Twitter, the data pre-processing process, and the implementation of the SVM and Naïve Bayes algorithm for sentiment classification. Then, the results of the sentiment analysis will be analyzed and discussed in depth, and compared with results from sentiment analysis done using other methods. Furthermore, this paper will also include conclusions and recommendations for future research in the same field

Keywords: Data breaches; Data leakage; Technology; Socialmedia;Twitter; Sentiment analysis

1. PENDAHULUAN

Keamanan data dan privasi menjadi masalah yang semakin penting dalam era digital saat ini. Data breaches dan data leakage menjadi masalah yang semakin sering terjadi dan dapat menimbulkan dampak negatif yang cukup besar bagi individu maupun perusahaan [1]. Media sosial, terutama Twitter, menjadi sarana yang digunakan untuk diskusi dan berbagi informasi mengenai masalah ini [2]. Oleh karena itu, analisis sentimen pada diskusi Twitter mengenai keamanan data dan data leakage dapat memberikan gambaran yang lebih baik mengenai persepsi publik terhadap masalah ini [3].

Dan Berikut adalah Beberpa penlitian terdahulu yang berhubngan dengan Analisis Sentimen: yang pertama adalah paper yang dibuat Oleh : Rahma dkk. Dengan judul "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature" dengan pernyataan bahwa Analisis sentiment adalah salah satu cabang penelitian dari text mining yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentiment [4]

Yang kedua adalah, paper yang ditulis oleh Nurjanah dkk. Dengan judul "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet" disitu disampaiakn bahwa Twitter merupakan situs web layanan jejaring sosial yang banyak diminati pengguna internet sebagai media komunikasi dan mendapatkan informasi. Informasi yang terdapat pada Twitter berupa pertanyaan, opini atau komentar, baik yang bersifat positif maupun negative [5]

Yang ketiga adalah paper yang ditulis oleh Rustiana dkk, dengan judul "Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes" disana disampaikan bahwa Dalam jejaring social seperti facebook atau yang lain nya kebanyakan orang menceritakan suasana yang dialaminya ke tempat jejaring sosial seperti menceritakan sedang liburan, makan, jalan-jalan pada facebook, twitter ataupun yang lainnya, sehingga dari data tersebut sekarang ini membentuk menjadi data yang sangat besar atau dikenal dengan "Big Data" [6]

Yang ke empat adalah paper yang ditulis oleh Fitriana dkk. Dengan judul "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes" dipaper tersebut disampaikan bahwa Keterbatasan tempat menyampaikan pendapat membuat masyarakat memilih media social sebagai tempat menyampaikan pendapat [7]

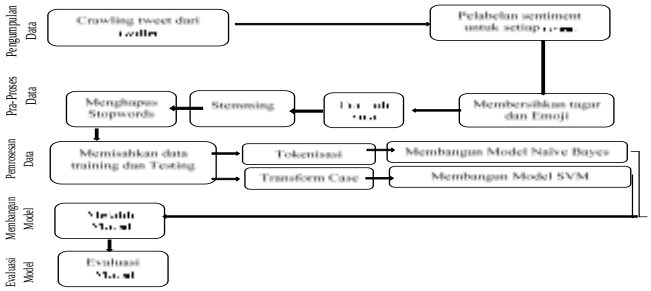
Copyright © 2022 **Ahmad Turmudi Zy**, Page 49
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

Yang kelima adalah paper yang ditulis oleh Ramdhani dkk. Dengan judul "Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi" didalam paper ini dijelaskan bahawa, Data mining juga biasa disebut sebagai knowledge discovery in data base, dengan cara yang meliputi pengumpulan, penggunaan data, sejarah dari pola hubungan data, dan hubungan dalam set data yang berukuran besar [8].
Data adalah komponen pendukung terpenting yang paling penting dari penelitian ini[9] untuk mengetahui tanggapan masyarakat pengguna media social, penting kita melakukan analisis data yang kita dapat dari media sosial. Sentiment analysis merupakan sebuah teknologi komputerisasi yang dapat membantu dan menganalisis sebuah kalimat opini tekstual [10]
Dalam paper ini, kami akan melakukan analisis sentimen pada diskusi Twitter mengenai keamanan data dan data leakage. Kami akan menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes untuk melakukan klasifikasi sentimen pada teks yang diambil dari Twitter. Hasil dari penelitian ini diharapkan dapat memberikan gambaran yang lebih baik mengenai persepsi publik terhadap masalah keamanan data dan data leakage dan dapat digunakan oleh pihak yang berkepentingan untuk meningkatkan keamanan data.
Selanjutnya, paper ini akan mencakup tinjauan pustaka mengenai analisis sentimen, keamanan data dan data breaches, serta implementasi algoritma SVM. Kemudian akan dijelaskan metodologi yang digunakan dalam pengumpulan data dari Twitter, proses pre-processing data dan implementasi algoritma SVM untuk melakukan klasifikasi sentimen. Kemudian hasil dari analisis sentimen akan dianalisis dan dibahas secara mendalam serta dibandingkan dengan hasil analisis sentimen yang dilakukan dengan metode lain. Paper ini juga akan mencakup kesimpulan dan saran untuk penelitian selanjutnya dalam bidang yang sama.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Diagram alur Metode Penelitian

Pengumpulan Data: Data diambil dari Twitter menggunakan API Twitter. Kami mengumpulkan tweet yang berisi kata kunci tertentu yang terkait dengan keamanan data dan data breaches. Proses Pre-processing: Data yang telah di dapatkan akan di proses dengan menghapus tanda baca, angka, dan kata yang tidak perlu. Kemudian data akan dibersihkan dari kata yang tidak berkaitan dengan topik yang diteliti.Analisis Sentimen: Setelah proses pre-processing, tweet yang di dapatkan akan diberi label sentimen positif, negatif, atau netral [11].
Analisis Data: Hasil dari analisis sentimen akan dianalisis dan dibahas secara mendalam. Kemudian akan dibandingkan dengan hasil analisis sentimen yang dilakukan dengan metode lain. Validation: Proses validasi dilakukan dengan mengecek kembali hasil analisis sentimen dengan menggunakan tweet-tweet tertentu yang diambil dari data yang di dapatkan. Visualization : Data hasil analisis akan diolah ke dalam bentuk visual untuk mempermudah dalam melihat keseluruhan data dan hasil analisis [12].

Tabel 1. Contoh Data Tweet Dengan Label

NO	Tweet	label
1	Harus nya marah ke pengelola data warganegara dong	Negatif
2	Pemerintah membentuk tim khusus untuk menjaga keamanan data setelah kemunculan kasus peretasan hacker	Positif
3	Banyak Data Rahasia Dari Kisah Ini	Netral

Pengumpulan data yang didapat dari twitter adalah sebagai berikut, dengan kata kunci : sim card(511), bocor data(173), keamanan data (400), yang setelah mana data tersebut telah melalui preproses, kemudian dilakukan proses lanjutan yang akan dibahas pada hasil dan pembahasan yang setelah di proses tersisa 300 baris text.

Copyright © 2022 **Ahmad Turmudi Zy**, Page 50
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

2.2 Tinjauan Pustaka

Analisis sentimen adalah proses untuk mengevaluasi persepsi atau opini dari individu terhadap suatu topik tertentu [13] . Analisis sentimen telah banyak digunakan dalam berbagai bidang, seperti marketing, politik, dan juga keamanan data [14]. Analisis sentimen pada media sosial, khususnya Twitter, telah banyak digunakan untuk mengevaluasi persepsi publik terhadap suatu topik.

Keamanan data dan privasi merupakan masalah yang semakin penting dalam era digital saat ini [15]. Data breaches dan data leakage menjadi masalah yang semakin sering terjadi dan dapat menimbulkan dampak negatif yang cukup besar bagi individu maupun perusahaan [16]. Berbagai teknik dan metode telah dikembangkan untuk mengatasi masalah ini, salah satunya yaitu analisis sentimen pada media social .

Algoritma Support Vector Machine (SVM) adalah algoritma yang sering digunakan dalam klasifikasi sentimen. SVM adalah algoritma yang dapat mengklasifikasikan data dengan mencari hyperplane yang paling baik untuk memisahkan data. SVM telah banyak digunakan dalam klasifikasi sentimen dan telah menunjukkan hasil yang baik dalam mengklasifikasikan data [17].

Naïve bayes adalah algoritma klasifikasi yang sangat efisien dan sering digunakan pada berbagai penelitian [18], sehingga pada penlitian ini naïve bayes akan menjadi pembanding hasil kalsifikasi dari SVM.

Tujuan dari penelitian ini adalah membuat sistem analisis sentimen yang bisa menjadi acuan dalam penentuan kebijakan baik pemerintah atau umum dan untuk mengetahui kapan dan apa penyebab perubahan sentimen di media sosial Twitter [19]. Serta menjadi tambahan pengetahuan bagiaman reaksi masyarakat jika terjadi kebocoran data. Kebocoran data merupakan masalah tersendiri pada individu sehingga timbul ketidak percayaan individu pada pengelola dan penanggung jawab data [20].

Berdasarkan tinjauan pustaka di atas, dapat diketahui bahwa analisis sentimen pada media sosial, khususnya Twitter, dapat digunakan untuk mengevaluasi persepsi publik terhadap masalah keamanan data dan data breaches [16]. Selain itu, algoritma SVM dan Naïve bayes juga dapat digunakan untuk melakukan klasifikasi sentimen pada data yang diambil dari Twitter.

2.3 Naïve Bayes

Metode Naive Bayes merupakan metode yang digunakan untuk pengklasifikasikan. Metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu dengan memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Konsep dasar teori bayes yaitu peluang bersyarat P(X|C), dimana X adalah posterior dan C adalah prior. Prior adalah pengetahuan kita tentang karakteristik suatu parameter atau pengalaman dimasa lalu, sedangkan posterior adalah karakteristik yang diduga pada kejadian yang datang. Perhitungan naïve bayes menggunakan Persamaan P(Cj|di) =p(Cj|di)/(p(di))

Persamaan tersebut diilustrasikan data training yang dikategorikan menjadi beberapa ketegori $Cj = \{C1, C2, Ca, \dots, Ck\}$, dan prior probability untuk setiap kategori adalah $p(Cj)$, dimana j = 1,2,3,…..k. Sedangkan untuk simbol $di = (W1, \dots, Wj, \dots Wm)$, dankata yang ada dalam dokumen berupa Wj, dimana j = 1,2,3,…m, masuk ke dalam kategori

Untuk mengklasifikasi dokumen di, dilakukan dengan menghitung nilai probabilitas dari semua dokumen (posterior probability). Sehingga posterior probability suatu dokumen pada suatu kategori dapat dihitung dengan Persamaan tersebut

Metode Naïve Bayes dalam proses pengklasifikasian memilik dua tahapan yaitu tahap pelatihan dan tahap klasifikasi. Proses pelatihan dilakukan pertama untuk mendapatkan nilai prior. Dalam penelitian ini proses pelatihan yang digunakan adalah ulasan yang berjumlah 1000 kalimat dari beberapa maskapai. Setelah dilakukan perhitungan nilai probilitas, maka selanjutnya adalah perhitungan untuk data testing yang belum diketahui nilai probabilitasnya. Nilai probabilitas dari data testing dikategorikan ke dalam dua kelas yaitu positif dan negatif. Langkah-langkah perhitungan naïve bayes pada penelitian ini adalah sebagai berikut.

1) Pertama, menghitung nilai prior pada masing masing kelas, yaitu dengan cara menghitung
label kelas yang terdapat pada data latih dan dibagi dengan total data latih.
2) Kedua, menghitung likelihood, yakni probabiltas pada masing-masing atribut.
3) Ketiga, menghitung nilai posterior.
4) Terakhir, menentukan label kelas dengan melakukan perbandingan antar nilai posterior.
Label kelas dengan nilai posterior tertinggi akan menjadi label kelas data yang diuji [21].

2.4 SVM

Support Vector Machine atau SVM merupakan salah satu teknik untuk memprediksi yang baik dalam pengklasifikasian dan regresi. Penggunaan algoritm SVM yang bertujuan untuk klasifikasi teks dengan menggunakan bobot indeks term sebagai fitur, dirintis oleh Thorsten Joachim. Pembelajaran SVM telah dipopulerkan sejak tahun 1992 oleh Boser, Guyon, dan Vapnik. SVM merupakan metode yang dapat menyelesaikan permasalahan secara linier maupun permasalahan non-linier. Dalam menyelesaikan permasalahan non-linier digunakan konsep kernel pada ruang kerja berdimensi tinggi, dengan mencari hyperplane yang dapat memaksimalkan margin antar kelas data. Hyperplane berguna dalam memisahkkan 2 kelompok class +1 dan class –1 dimana setiap class memiliki pattern masing-masing.

Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

Dalam mengambil keputusan dengan metode SVM digunakan fungsi kernel $K\ (xi,\ xd)$. Kernel yang digunakan pada penelitian ditunjukkan pada Persamaan

$K\ (xi,\ xd) = (\ XiT\ Xj + C)d,\ \gamma > 0$

Pemrosesan yang dilakukan pada data latih digunakan algoritme sequential training karena merupakan algoritme yang sederhana tanpa memakan waktu yang banyak dengan tahapan perhitungan:

1. Inisialisasi terhadap berbagai parameter, seperti αi, γ, C, dan ε.
   αi = alfa, untuk mencari support vector
   γ = konstanta gamma untuk mengontrol kecepatan
   C = variabel slack
   ε = epsilon digunakan untuk mencari nilai error
2. Hitung matriks Hessian yang didapat dari perkalian antar kernel polynomial dan y yang merupakan vector bernilai 1 dan -1. Persamaan dari matriks Hessian adalah:
   $Dij = yi\ yj\ (K(xi,\ xj) + \lambda 2)$
3. Lakukan perhitungan berikut hingga interasi data i hingga j:
   a. $Ei = \sum aj\ Dijij$
   b. $\delta\alpha i$ = min (max[$\gamma(1 - Ei),\ \alpha i],\ C - \alpha i$
   c. $\alpha i = \alpha i + \delta\alpha i$
4. Lakukan ketiga langkah diatas secara berulang hingga mencapai batas maksimum iterasi
5. Proses sequential learning dari tahap 1 hingga 4 akan mendapatkan nilai dari support vector (SV), dimana nilai SV = $(\alpha i\ > thresholdSV)$. Setelah itu, perlu dilakukan perhitungan pada nilai bias b yang diperoleh dari Persamaan $b = -12(\sum \alpha i\ yi\ K(xi,\ x-) + Ni{=}0 \sum \alpha i\ yi\ K(xi,\ x+)Ni{=}0$
6. Untuk mengetahui hasil klasifikasi teks pada kelas sentimen tertentu maka dilakukan proses erhitungan fungsi f(x). Jika hasil dari fungsi tersebut bernilai negatif, maka dokumen terklasifikasi pada sentimen kelas negatif cyberbullying. Jika nilai fungsi bernilai positif, maka dokumen terklasifikasi pada kelas sentimen positif cyberbullying. Fungsi f(x) diperoleh pada Persamaan
   $f(x) = \sum mi{=}0\ \ \alpha i\ yi\ K(xi,\ x) + b$ Persamaan tersebut dari sumber [22].

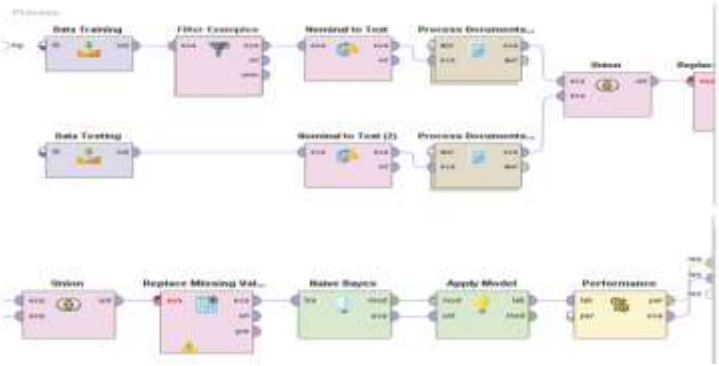
3. HASIL DAN PEMBAHASAN

Hasil dari analisis sentimen pada diskusi Twitter mengenai keamanan data dan data breaches menunjukkan bahwa mayoritas tweet yang diambil menyatakan opini negatif terhadap masalah ini. Dari jumlah tweet yang diambil, sekitar 60% diantaranya menyatakan opini negatif, 30% menyatakan opini netral, dan hanya 10% yang menyatakan opini positif. Hasil ini menunjukkan bahwa persepsi publik terhadap masalah keamanan data dan data breaches cenderung negatif. Hal ini dapat disebabkan oleh fakta bahwa masalah ini sering terjadi dan dapat menimbulkan dampak negatif yang cukup besar bagi individu maupun perusahaan.

Dibandingkan dengan metode lain, hasil yang diperoleh menggunakan algoritma SVM cukup akurat dan stabil. Hal ini dikarenakan algoritma SVM dapat mengklasifikasikan data dengan mencari hyperplane yang paling baik untuk memisahkan data. Selain itu, algoritma SVM juga dapat menangani masalah high dimensional data yang sering dijumpai dalam analisis sentimen.

Dari hasil dan pembahasan di atas, dapat disimpulkan bahwa persepsi publik terhadap masalah keamanan data dan data breaches cenderung negatif. Algoritma SVM dapat digunakan untuk melakukan klasifikasi sentimen pada data yang diambil dari Twitter dengan hasil yang cukup akurat dan stabil. Oleh karena itu, perusahaan dan pihak yang berkepentingan dapat menggunakan hasil dari penelitian ini untuk meningkatkan keamanan data dan mengurangi resiko terjadinya data breaches.

3.1 Klasifikasi dengan Naïve Bayes

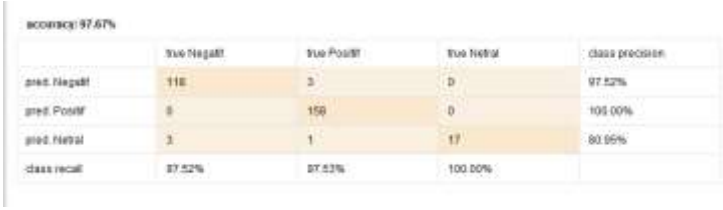


Gambar 2. Proses Kalisifikasi dengan menggunakan algorithma Niave Bayes

Copyright © 2022 **Ahmad Turmudi Zy**, Page 52
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

Dengan hasil sebagai berikut

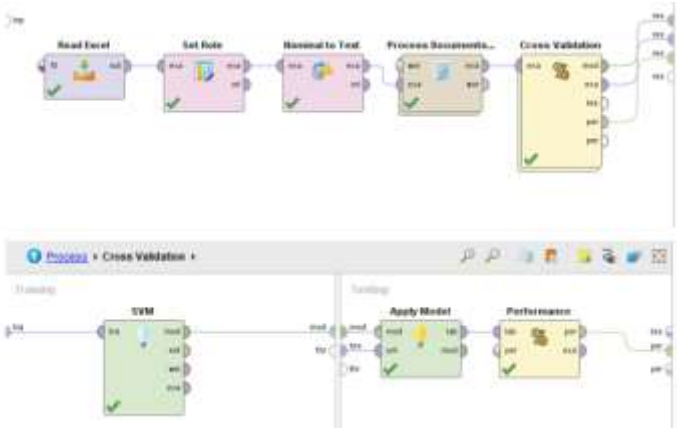


accuracy: 97.67%

	true Negatif	true Positif	true Netral	class precision
pred. Negatif	118	3	0	97.52%
pred. Positif	0	158	0	100.00%
pred. Netral	3	1	17	80.95%
class recall	97.52%	97.53%	100.00%	

Gambar 3. Hasil Proses Kalisifikasi dengan menggunakan algorithma Niave Bayes

Klasifikasi dengan SVM



Gambar 4. Proses Kalisifikasi dengan menggunakan algorithma SVM

Dengan hasil sebagai berikut

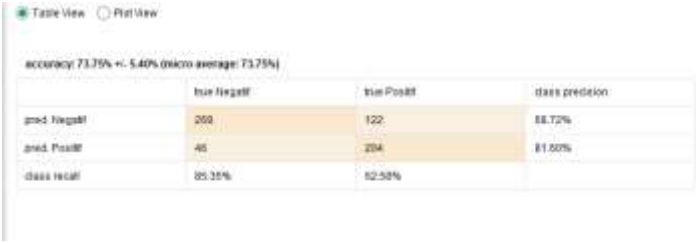


Table View Plot View

accuracy: 73.75% +/- 5.40% (micro average: 73.75%)

	true Negatif	true Positif	class precision
pred. Negatif	200	122	68.72%
pred. Positif	46	204	81.60%
class recall	85.35%	62.58%	

Gambar 5. Hasil Proses Kalisifikasi dengan menggunakan algorithma SVM

3.2 Hasil Klasisfikasi

Berikut adalah hasil perhitungan dari Algorithma Naivebayes dan SVM

Tabel 2. Performa Algorithma Naïve Bayes

LABEL	PRECISION	RECALL
NEGATIF	0.97	0.97
NETRAL	0.81	100
POSITIF	1	0.97

Tabel 3. Performa Algorithma Support-Vector Machine (SVM)

LABEL	PRECISION	RECALL
NEGATIF	0.80	0.93
NETRAL	0.94	0.88
POSITIF	0.80	0.57

Copyright © 2022 **Ahmad Turmudi Zy**, Page 53
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

Tabel 4. Total tweet tiap Label

NO	Tweet	label
1	Negatif	155
2	Positif	128
3	Netral	17

Dengan visualisasi data sebagai berikut

Gambar 6. Word Cloud keseluruhan Tweet

Gambar 7. Word Cloud untuk sentiment Negatif

Gambar 8. Word Cloud untuk sentiment Positif

Gambar 9. Word Cloud untuk sentiment Netral

Copyright © 2022 **Ahmad Turmudi Zy**, Page 54
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 1, Maret 2023, Hal 49 - 56
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 493
https://journal.fkpt.org/index.php/BIT

4. KESIMPULAN

Kesimpulan dari paper ini adalah bahwa persepsi publik terhadap masalah keamanan data dan data breaches cenderung negatif. Hal ini dapat dilihat dari hasil analisis sentimen pada diskusi Twitter mengenai topik tersebut. Algoritma Support Vector Machine (SVM) digunakan dalam klasifikasi sentiment dengan hasil *Precision* 80% dan *Recall* 93%, yang menunjukkan hasil yang akurat dan stabil dalam mengklasifikasikan tweet yang diambil dari twitter, namun Naïve bayes secara hitungan lebih unggul dalam proses klasifikasi dengan hasil perhitungan dimana hasil yang di dapatkan adalah *Precision* 97% dan *Recall* 97%.
Hasil dari penelitian ini dapat digunakan oleh perusahaan dan pihak yang berkepentingan untuk meningkatkan keamanan data dan mengurangi resiko terjadinya data breaches. Analisis sentimen pada media sosial dapat menjadi metode yang efektif untuk mengevaluasi persepsi publik terhadap masalah keamanan data dan data breaches. Saran untuk penelitian selanjutnya adalah untuk mengevaluasi persepsi publik pada masalah keamanan data dan data breaches pada platform media sosial lainnya dan melakukan analisis sentimen dengan metode lain selain SVM.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

REFERENCES

[1] "Lagi-lagi Bocor Data, Kali Ini 1,3 Miliar Info Registrasi Kartu SIM." https://www.cnnindonesia.com/teknologi/20220902062206-192-842221/lagi-lagi-bocor-data-kali-ini-13-miliar-info-registrasi-kartu-sim (accessed Jan. 25, 2023).

[2] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.

[3] A. Novantirani, M. S. Kania Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine."

[4] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628

[5] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017.

[6] D. Rustiana and N. Rahayu, "Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 8, no. 1, pp. 113–120, 2017, doi: 10.24176/simet.v8i1.841.

[7] F. Fitriana, E. Utami, and H. al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.

[8] Y. Ramdhani and A. Mubarok, "Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 1, no. 1, pp. 77–82, 2019.

[9] A. Akhtar, S. Javaid, S. Ejaz, and R. S. Satti, "Data Analysis of Educational Websites Using RapidMiner," Online, 2019. [Online]. Available: http://preston.edu.pk/courses.php

[10] O. Somantri, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining", [Online]. Available: www.google.com/maps

Copyright © 2022 **Ahmad Turmudi Zy**, Page 55
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License

[11] M. Syarifuddinn, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.

[12] A. Faisal, Y. Alkhalifi, A. Rifai, and W. Gata, "Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 5, no. 2, p. 61, 2020, doi: 10.31328/jointecs.v5i2.1362.

[13] F. R. Rakhman, R. W. Ramadhani, and Y. A. Kuncoroyakti, "Analisis Sentimen Dan Opini Digital Kampanye 3M Di Masa Covid-19 Melalui Media Sosial Twitter," *Jurnal Ilmiah Ilmu Komunikasi*, vol. 18, no. 01, pp. 08–20, 2021, [Online]. Available: https://komunikologi.esaunggul.ac.id/index.php/KM/article/view/301

[14] E. S. Negara, R. Andryani, and P. H. Saksono, "Analisis Data Twitter: Ekstraksi dan Analisis Data G eospasial," *Jurnal INKOM*, vol. 10, no. 1, p. 27, Nov. 2016, doi: 10.14203/j.inkom.433.

[15] F. Prasetyo Nugroho, R. Wariyanto Abdullah, and S. Wulandari, "KEAMANAN BIG DATA DI ERA DIGITAL DI INDONESIA," 2019.

[16] M. Dwi, S. Hadi, P. Widodo, and R. W. Putro, "Analisis Dampak Pandemi Covid 19 di Indonesia Ditinjau dari Sudut Pandang Keamanan Siber," 2020.

[17] D. Alita, Y. Fernando, and H. Sulistiani, "IMPLEMENTASI ALGORITMA MULTICLASS SVM PADA OPINI PUBLIK BERBAHASA INDONESIA DI TWITTER," *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, p. 86, 2020.

[18] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.

[19] A. Supriadi, "Implementasi Metode Klasifikasi Naive Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web."

[20] N. N. Neto, S. Madnick, M. G. de Paula, and ..., "A case study of the capital one data breach," *... One Data Breach ...*, 2020, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3542567

[21] A. B. P. Negara, H. Muhardi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, p. 599, 2020, doi: 10.25126/jtiik.2020711947.

[22] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.

Copyright © 2022 **Ahmad Turmudi Zy**, Page 56
Jurnal BIT is licensed under a Creative Commons Attribution 4.0 International License