

TUGAS AKHIR

**ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP
KASUS KEBOCORAN DATA MASYARAKAT INDONESIA
MENGUNAKAN ALGORITMA *SUPPORT VECTOR
MACHINE (SVM)***



Disusun Oleh:

Teo Yulio Sihotang 1923250052

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN REKAYASA
UNIVERSITAS MULTI DATA PALEMBANG
PALEMBANG
2023**

**Fakultas Ilmu Komputer dan Rekayasa
Universitas Multi Data Palembang**

Program Studi Informatika
Tugas Akhir Sarjana Komputer
Semester Gasal Tahun 2022/2023

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KASUS
KEBOCORAN DATA MASYARAKAT INDONESIA MENGGUNAKAN
*SUPPORT VECTOR MACHINE (SVM)***

Teo Yulio Sihotang

1923250052

Abstrak

Kebocoran Data merupakan sebuah kejadian yang terjadi dikarenakan ada kerenggangan terhadap sistem keamanan komputer yang mengganggu data-data private yang tidak bisa dijangkau orang lain. Indonesia 1 tahun kebelakang sudah terjadi insiden kebocoran data yang seharusnya tidak terjadi pada instansi-instansi besar. Maka dari kejadian ini dilakukan penelitian menggunakan algoritma *support vector machine* dengan menggunakan data yang diambil melalui proses crawling dengan total keseluruhan 1000 data tweet berdasarkan hastag “#databocor”, “#kebocorandata”. pengujian dibuat menjadi 3 skenario yang berbeda dengan perbandingan pengujian dari pembagian data 70:30, 60:40 dan 50:50. Pada pengujian keseluruhan skenario klasifikasi mendapatkan hasil akurasi rata-rata sebesar 95.75%, presicion sebesar 93.5, recall sebesar 100% dan untuk pengujian keseluruhan skenario validasi mendapatkan hasil rata-rata accuracy sebesar 78.33%. Pada keseluruhan data setelah dilakukan pelabelan perbandingan sentimen positif memiliki persentase sebesar 74% dan sentimen negatif memiliki persentase sebesar 26%. Dari hasil yang didapatkan sentimen positif memiliki persentase lebih besar dibanding sentimen negatif.

Kata Kunci: Data Bocor, Kebocoran Data, *Support Vector Machine*, Analisis Sentimen



BAB 1

PENDAHULUAN

Bab ini merupakan pembahasan tentang latar belakang, rumusan masalah, ruang lingkup, tujuan, manfaat, metodologi penelitian dan sistematika penulisan yang dilakukan peneliti.

1.1 Latar Belakang Masalah

Kemajuan teknologi yang terjadi saat ini cukup cepat dirasakan oleh masyarakat dalam kehidupan sehari-hari. Kemajuan teknologi ini beriringan dengan penggunaan internet yang dapat membantu memudahkan kegiatan masyarakat di seluruh dunia. teknologi dan informasi sebagai salah satu alat bantu yang digunakan manusia dalam melakukan pekerjaan (Budiman, 2019). Dengan bantuan teknologi dan disertai kemajuaan internet yang sudah semakin memadai banyak hal yang bisa dilakukan dalam membantu kegiatan kehidupan manusia agar menjadi lebih cepat dan efisien, contohnya seperti *whatsapp*, *line*, *facebook*, *instagram*, *twitter* dan aplikasi lainnya. hanya dengan sebuah benda kecil yang canggih yaitu *smartphone* kita bisa terhubung dengan orang yang jauh dalam waktu yang lama.

Media sosial adalah hasil dari perkembangan kemajuan teknologi yang yang dapat dirasakan oleh masyarakat pada saat ini. Media sosial adalah salah satu contoh

penghasil data internet terbesar (Dhawan, 2014), Banyak sekali keuntungan bagi masyarakat menggunakan media sosial seperti mencari teman baru, mencari

informasi update dan menjadikan wadah untuk menuangkan sebuah opini ataupun kritik pada suatu objek yang sedang viral. Lebih dari jutaan orang atau organisasi menggunakan media sosial sebagai bentuk eksistensinya di dunia maya (Sivarajah, 2019). Twitter merupakan salah satu contoh *platform* yang digunakan pengguna untuk menuangkan isi pikiran melalui tulisan *online* yang berkaitan dengan kejadian yang sedang terjadi ataupun tidak terjadi.

Kemajuan teknologi seharusnya beriringan dengan kemajuan keamanan data. banyak ancaman yang terjadi dengan kemajuan teknologi yang dapat merugikan pengguna jika tidak ada sistem keamanan yang terbaru mengatasi ancaman yang akan terjadi pada kemajuan teknologi seperti contoh *phising*, pencurian akun, kebocoran data, dan tindakan sejenisnya. Kebocoran data merupakan hal yang tidak diinginkan semua orang yang dimana banyak hal yang dapat dilakukan dengan data tersebut untuk merugikan orang lain dan hal ini tidak lain hanya ingin untuk memperkaya diri ataupun hal lain yang menjadi pertimbangan dengan membocorkan data tersebut. Dilansir dari (Suara.com. 2022) bahwasanya Indonesia sekarang sedang mengalami darurat keamanan siber yang dimana pada bulan agustus kemarin dalam satu bulan sudah mengalami kebocoran data sebanyak 3 kali berawal dari 26 Juta data pelanggan IndiHome, 17 juta data pengguna PLN hingga 1.3 Miliar Data pengguna kartu *Subscriber identity module* (SIM).

Semua kejadian berkaitan dengan data diri pribadi seperti Nomor Induk Kependudukan (NIK), Nomor telepon, Nama sampai alamat tempat tinggal. Banyak ketakutan dan keraguan masyarakat timbul dikarenakan terjadinya kebocoran data ini, jika data yang bocor di tangan yang tidak bertanggung jawab maka yang ditakutkan adalah data yang ada disalah gunakan untuk hal-hal yang merugikan masyarakat.

Pada penelitian yang dilakukan Muh. Fitra Rizki, Karina Auliasari, Renaldi Primaswara Prasetya (2021) mengenai analisis sentiment *cyberbullying* pada sosial media twitter menggunakan metode *support vector machine*. Dari penelitian yang dilakukan dapat dilihat hasil akhir yang didapatkan pada penggunaan metode *Support Vector Machine* telah mampu mengklasifikasi dengan baik dengan catatan akurasi pada percobaan 100 data dokumen tweet menghasilkan nilai recall 64%, precision 58% dan tingkat akurasi sebesar 70 % dan juga dalam percobaan menggunakan 200 penggunaan data dokumen tweet mendapatkan catatan akurasi pada nilai *recall* 71%, *precision* 63% dan tingkat akurasi sebesar 75%.

Pada penelitian yang dilakukan Primandani Arsi, Retno Waluyo (2020) mengenai analisis sentiment wacana pemindahan ibu kota Indonesia menggunakan algoritma *support vector machine* (SVM). Dari penelitian yang dilakukan dapat dilihat hasil akhir yang didapatkan pada penggunaan algoritma support vector machine dengan proses pengujian telah dilakukan peneliti terkait pemindahan ibu kota terdapat total tweet 1.236 dikelompokkan 404 positif dan 832 negatif

menggunakan metode SVM mendapatkan hasil yaitu akurasi 96,68 %, *precision* 95,82%,*recall* 94,04%.

Pada penelitian yang dilakukan Bergas Pamungkas, Muhammad Eka Parbaya, Dwi Januarita A.k (2021) mengenai analisis sentiment Twitter menggunakan *Support Vector Machine*(SVM) pada kasus benih lobster 2020. Dari penelitian yang dilakukan dapat dilihat hasil akhir yang didapatkan pada penggunaan algoritma SVM terdapat hasil 28% dalam kategori emosi negatif dan 72 dari emosi positif dan dari keseluruhan hasil terdapat nilai akurasi sebesar 84,21%, nilai sensitivitas 73,38 % dan nilai spesifisitas 82,10%.

Pada penelitian yang dilakukan Lutfanida (2022) mengenai analisis sentiment data twitter menggunakan metode naïve bayes dan *support vector machine* (SVM) tentang presiden Jokowi 3 periode. Dari hasil penelitian ini dapat diperoleh kesimpulan yaitu dari metode *Support Vector Machine* mendapatkan hasil yaitu tingkat akurasi sebesar 95,42%, *precision negative* sebesar 98,01% dan positif sebesar 93,09%. Dan algoritma naïve bayes sebesar 94,07%. Dari hasil penelitian ini dapat diperoleh kesimpulan bahwa Algoritma *support vector machine* (SVM) lebih tinggi dibandingkan *naïve bayes*.

Pada penelitian yang dilakukan Hendrik Setiawan, Ema Utami, Sudarmawan (2022) mengenai analisis sentiment twitter kuliah online pasca Covid-19 menggunakan algoritma *support vector machine* dan naïve Dari penelitian yang dilakukan dapat dilihat hasil akhir yang didapatkan pada penggunaan algoritma hasil

terbaik yang dicapai algoritma naïve bayes dalam proses pembobotan mendapatkan akurasi sebesar 81,20% dengan waktu 9,00 detik, recall 79,60% dan presisi 79,40% dan untuk hasil terbaik dari algoritma *support vector machine* yaitu masuk kedalam iterasi ke 423 yaitu dalam proses training data pada algoritma *support vector machine* mendapatkan hasil akurasi sebesar 85% waktu 31,60 detik, recall 84% dan presisi 83,60 dengan ini algoritma *support vector machine* mendapatkan akurasi lebih tinggi dibandingkan algoritma naïve bayes.

Pada penelitian yang dilakukan Mohammad Rezza Fahlevi (2022) mengenai Analisis sentimen terhadap ulasan aplikasi pejabat pengelola informasi dan dokumentasi kementerian dalam negeri republic Indonesia di google playstore menggunakan metode support vector machine. Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa support vector machine dengan data berasal dari google playstore berjumlah 700 data yang dibagi menjadi 85 positif dan 615 negatif dalam tahapan evaluasi mendapatkan hasil analisis sentimen akurasi sebesar 97%, precision 94% dan recall 100%.

Pada penelitian yang dilakukan Rizki Maulana, Sri Redjeki (2016) mengenai analisis sentimen pengguna twitter menggunakan metode support vector machine berbasis cloud computing. Dari penelitian yang dilakukan dapat dilihat hasil akhir yang didapatkan pada penggunaan algoritma support vector machine dalam penelitian ini mendapatkan hasil 79,5% untuk akurasi, hasil ini dinyatakan cukup baik oleh peneliti dikarenakan mengingat objek yang diteliti ini adalah kalimat

bahasa Indonesia dengan struktur dan variasi kalimat yang berbeda dan bahasa yang lebih rumit.

Berdasarkan fenomena dan beberapa hasil penelitian di atas yang berbeda-beda, maka dari itu diperoleh judul untuk penelitian ini yaitu Analisis Sentimen Pengguna Twitter Terhadap Kasus Kebocoran Data Menggunakan Algoritma *Support Vector Machine* (SVM), untuk mengetahui tanggapan pengguna twitter terkait topik kebocoran data yang sedang terjadi. Hasil akhir dapat menjadi bahan edukasi dan evaluasi instansi terkait peningkatan keamanan siber demi keamanan data masyarakat.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian adalah bagaimana sentimen pengguna twitter terhadap kasus kebocoran data masyarakat Indonesia yang diukur dengan menggunakan algoritma *Support Vector Machine*?

1.3 Ruang Lingkup

Ruang lingkup dari penelitian analisis sentimen sebagai berikut:

1. Algoritma yang digunakan adalah algoritma Support Vector Machine (SVM).
2. Dataset penelitian adalah Tweet pengguna pada media sosial Twitter
3. Penelitian ini berfokus pada sentimen tweet pengguna media sosial

Twitter

4. Dataset yang telah dikumpulkan melalui proses crawling dengan bantuan library API Twitter.
5. Dataset yang digunakan adalah dataset yang telah dikumpulkan mulai dari (1 Agustus 2022 – 1 Oktober 2022) dengan kata kunci yaitu ‘#databocor’, ‘#kebocorandata’.
6. Dataset yang digunakan dalam penelitian analisis sentimen ini berjumlah 1000 data yang dimana telah dibagi menjadi 70% data latih dan 30% data uji melalui proses splitting data.
7. Pemberian label dataset dilakukan manual oleh bantuan ibu Tiasa Singalingging S.pd sebagai Guru Bahasa Indonesia di SD N 20 Tanjung Enim.
8. Pelabelan data dibagi menjadi 2 kelompok yaitu label positif dan label negatif
9. Penggunaan library NLTK dan Sastrawi Bahasa Indonesia yang digunakan sebagai alat bantu pada penelitian dalam proses preprocessing.
10. Penggunaan tools Google Colab dan Bahasa pemrograman Python dalam proses penelitian yang dilakukan oleh peneliti.

1.4 TUJUAN DAN MANFAAT

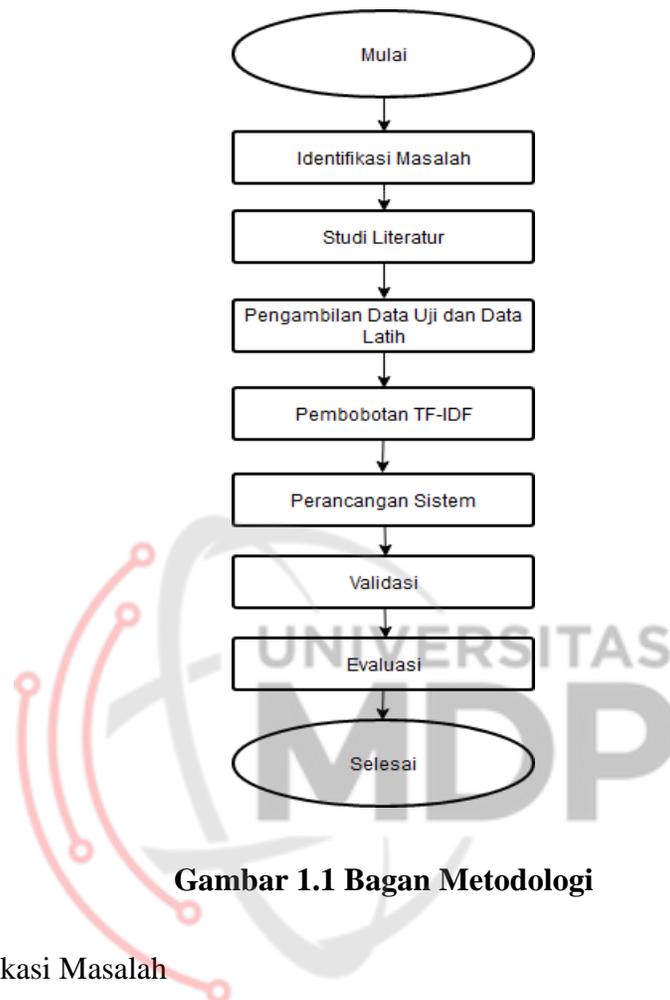
Tujuan penelitian ini adalah untuk mengetahui respon dari pengguna twitter terhadap kasus kebocoran data yang lagi terjadi pada saat ini berdasarkan tweet yang dilakukan oleh pengguna media sosial Twitter dengan menggunakan Algoritma *Support Vector Machine (SVM)*.

Manfaat dari penelitian ini adalah:

1. Memahami Bagaimana proses pengumpulan data untuk melakukan kegiatan penelitian.
2. Memahami cara menggunakan dan menerapkan Algoritma support vector machine (SVM).
3. Mengetahui perbandingan sentimen positif dan sentimen negatif dalam proses kesimpulan akhir penelitian.
4. Mempermudah pemerintah untuk mengetahui informasi keresahan masyarakat Indonesia terhadap kebocoran data yang sudah terjadi.
5. Menambah hasil penelitian terkait dengan Algoritma Support Vector Machine (SVM).

1.5 METODOLOGI PENELITIAN

Berikut ini merupakan tahapan yang dilakukan peneliti untuk melakukan analisis sentimen pengguna twitter terhadap kasus kebocoran data masyarakat Indonesia. Bagan metodologi penelitian terdapat pada gambar 1.1 berikut



Gambar 1.1 Bagan Metodologi

1. Identifikasi Masalah

Tahapan identifikasi merupakan sebuah proses tahapan yang dilakukan oleh peneliti untuk mengetahui serta mengidentifikasi masalah mengenai analisis sentimen pengguna twitter terhadap kasus kebocoran data masyarakat Indonesia menggunakan algoritma support vector machine (SVM)

2. Studi Literatur

Studi literatur merupakan proses tahapan yang dilakukan oleh peneliti untuk melakukan pengumpulan informasi serta pemahaman terkait dengan jurnal-

jurnal yang terkait dengan algoritma yang digunakan dalam penelitian analisis sentiment pengguna twitter terhadap kasus kebocoran data masyarakat Indonesia menggunakan algoritma support vector machine (SVM)

3. Pengambilan Data Uji dan Data Latih

Pengambilan Data Uji dan Data Latih merupakan sebuah proses tahapan yang dilakukan oleh peneliti sebelum memulai masuk dalam penelitian yang dimana dilakukan dengan menggunakan tools Google Colab dan pemrograman Python serta API twitter dalam proses crawling data dari twitter dengan jumlah 1000 data yang diambil berdasarkan kata kunci '#databocor', '#kebocorandata'. Dalam range waktu pengambilan data yaitu (1 agustus – 1 oktober 2022) dan setelah itu langsung dilakukan proses preprocessing dan terakhir adalah splitting data untuk data uji dan data latih.

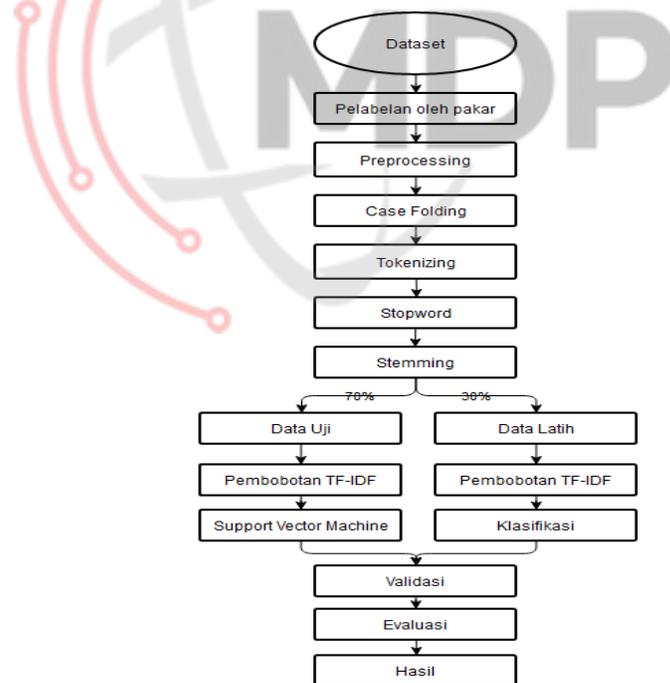
4. Pembobotan TF-IDF

Pembobotan TD-IDF merupakan sebuah proses yang dilakukan peneliti untuk merubah kata menjadi sebuah angka atau vektor yang nantinya akan di validasi seberapa banyak frekuensi kata itu muncul didalam sebuah dokumen.

5. Perancangan Sistem

Perancangan sistem merupakan sebuah tahapan lanjutan yang dilakukan peneliti setelah mendapatkan dataset yang kemudian akan dilakukan proses preprocessing yang dimana terdapat 1000 data tweet dari hasil crawling yang telah

dilakukan oleh pelabelan oleh ahli. Setelah dilakukan pelabelan maka selanjutnya akan dilakukan proses preprocessing selanjutnya seperti case folding, tokenizing, stopword dan stemming. Kemudian dilakukan splitting data yang dibagi menjadi 70% data latih dan 30% data, setelah splitting data akan dilakukan pembobotan TF-IDF untuk mengetahui seberapa besar frekuensi sebuah kata muncul pada sebuah dokumen dan terakhir adalah melakukan validasi dengan K-fold cross validation serta pengujian pada algoritma yang sudah ditentukan untuk mengetahui hasil akhir yang sudah dilakukan penelitian. Tahapan yang dilakukan peneliti dapat di lihat pada bagan perancangan sistem sebagai berikut.



Gambar 1.2 Bagan Perancangan Sistem

6. Validasi

Validasi merupakan sebuah tahapan yang dilakukan peneliti untuk melatih data yang digunakan dalam penelitian agar mengetahui hasil yang didapat dari sistem apakah sudah sesuai atau belum dengan yang diinginkan. Pada validasi ini menggunakan K- Fold Cross Validation yang dimana menggunakan nilai $k = 5$, karena penelitian rata-rata dalam melakukan validasi menggunakan nilai $k = 5$ dan nilai $k = 10$ (Hermawan, Yoannita, 2018).

7. Evaluasi

Evaluasi merupakan tahapan terakhir yang dilakukan oleh peneliti yang dimana pada setelah selesai melakukan pengujian dengan menggunakan algoritma yang dibutuhkan maka akan mendapatkan hasil confusion matrix dalam acuan perhitungan hasil akhir dari recall, precision dan accuracy.

1.6 SISTEMATIKA PENULISAN

Alur penulisan penelitian dapat dilihat sebagai berikut:

Bab 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, ruang lingkup, tujuan, manfaat, metode penelitian dan sistematika penulisan yang akan dilakukan oleh peneliti.

Bab 2 LANDASAN TEORI

Bab ini merupakan penjelasan tentang teori-teori serta jurnal sesuai dengan kebutuhan dalam analisis sentimen, algoritma support vector machine serta penelitian terdahulu.

Bab 3 METODE PENELITIAN

Bab ini merupakan penjelasan alur pengujian yang dilakukan oleh peneliti dalam melakukan penelitian

Bab 4 ANALISIS DAN PEMBAHASAN

Bab ini merupakan penjelasan hasil dari penelitian yang dilakukan oleh peneliti dengan menampilkan hasil akhir penelitian yang sudah didapatkan.

Bab 5 SIMPULAN DAN SARAN

Bab ini merupakan penjelasan tentang hasil akhir kesimpulan dari penelitian yang sudah dilakukan dan saran



DAFTAR PUSTAKA

- Bandjar, D. A., Warouw, M., & Marentek, A. (2019). Dampak Penggunaan Twitter Terhadap Pembelajaran Bahasa Inggris (Ditinjau Dari Persepsi Mahasiswa).
- Dhawan, Z. (2014). Big Data and Social Media . *Research Matters : A Cambridge*, 36-41.
- Feldman, R., & Sanger, J. (n.d.). The Text Mining Handbook : Advance Approaches to Analyzing Unstructured Data. In R. Feldman, & J. Sanger, *The Text Mining Handbook : Advance Approaches to Analyzing Unstructured Data*. (p. 424). Cambridge: Cambridge.
- Indriasti, X. J. (2022, Oktober 27). *Suara.com*. Retrieved from *Suara.com*: <https://www.suara.com/news/2022/09/02/115017/11-daftar-kasus-kebocoran-data-di-indonesia-sebulan-tiga-kali-kejadian>
- Irfani, F. F., Triyanto, M., Hartanto, A. D., & Kusnawi. (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine. *Jurnal Bisnis, Manajemen dan Informatik*, 258-266.
- Irsyad, H., & Pribadi, M. R. (2019). Klasifikasi Opini Masyarakat Terhadap Pertanian Indonesia dengan Naive Bayes pada Twitter. *Jurnal Teknik Informatika Unik St.Thomas (JTIUST)*. Volume 04 Nomor 1, 89-98.
- Praghakusuma, A. Z., & Charibaldi, N. (2021). Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantas Korupsi). *Jurnal Sarjana Teknik Informatika* , 33-42.
- Soen, G. I., Marlina, & Renny. (2022). Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants. *JITU : Journal Informatic Technology And Communication*, 25-30.
- Luthfanida. (2022). Analisis Sentimen Data Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM) Tentang Presiden Jokowi 3 Periode. *Journal of Information Technology Research*, 5-11.
- Nurjannah, M., Hamdani, & Astuti, I. F. (2013). Penerapan Algoritma Term Frequency

(TF-IDF) Untuk Text Mining. *Jurnal Informatika Mulawarman Vol. 8 No. 3 September 2013* , 110-113.

Pamungkas, B., Purbaya, M. E., & A.K, D. J. (2021). Analisis Sentimen Twitter Menggunakan Metode *Support Vector Machine* (SVM) pada Kasus Benih Lobster 2020. *Journal of Informatics, Information System, Software Engineering and Applications* , 11-20.

Rachmi, H., & Surniandari, A. (2020). Classification of Text Mining Review Oil Diffuser Products Using Naive Bayes Classification. *Jurnal Mantik Volume 4 Number 1 May 2020* , 187-192.

Rizki, M. F., Auliasari, K., & Prasetya, R. P. (2021). Analisis sentiment cyberbullying pada sosial media twitter menggunakan metode *support vector machine*. *JATI (Jurnal Mahasiswa Teknik Informatika)* , 548-556.

Setiawan, H., Utami, E., & Sudarmawan. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma *Support Vector Machine* dan NaiveBayes. *Jurnal Komtika (Komputasi dan Informatika)* , 43-51.

Sivarajah, U., Irani, Z., Gupta, S., & Mahroof, K. (2019, April). *Role of big data and social media analytics for business to business sustainability : A participatory web context industrial marketing management*.

Rani, S., & Bhatt, S. (2018). Sentiment Analysis on Twitter Data Using *Support Vector Machine*. *International Journal of Scientific Research in Computer Science Applications and Management Studies*, 365-370.

Sawaswati, N. W. (2013). Naive Bayes Classifier Dan *Support Vector Machines* Untuk Sentiment Analysis. *Seminar Nasional Sistem Informasi Indonesia*, 586-591.

Lin, Y., Wang, X., & Zhou, A. (2016). *Opinion Analysis For Online Reviews*. China: East China Normal University Scientific Reports.

Fahlevi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan dokumentasi kementerian dalam negeri republik indonesia di google playstore menggunakan metode *support vector machine*. *Jurnal Teknologi dan Komunikasi Pemerintahan* , 1-13.

Maulana, R., & Redjeki, S. (2016). Analisis Sentimen Pengguna Twitter Menggunakan Metode *Support Vector Machine* Berbasis Cloud Computing. *Jurnal TAM (Technology Acceptance Model)* , 23-28.