

Independent Campus Student Exchange Sentiment Analysis Using SVM

Putri Irhami¹, Eva Darnila^{2*}, Fadlisyah³

^{1,2,3} Universitas Malikussaleh, Indonesia

*Corresponding Author Email: eva.darnila@unimal.ac.id

ABSTRAK

Received: 16 January 2024
Revised: 24 January 2024
Accepted: 6 February 2024
Available online: 1 April 2024

Kata kunci : Support Vector Machine, pertukaran pelajar

Support Vector Machine (SVM) adalah metode pembelajaran mesin yang banyak digunakan untuk masalah regresi dan klasifikasi terutama klasifikasi review aplikasi. Pertukaran pelajar merupakan salah satu program yang harus dipersiapkan oleh perguruan tinggi. Program pertukaran pelajar dimaksudkan untuk mengurangi persoalan disparitas sarana dan prasarana pendidikan di Indonesia. Keunggulan dari pertukaran pelajar dapat mengatur waktu, memiliki kesadaran tinggi dalam berkomunikasi, mampu mengaku ketika dirinya mengalami permasalahan dan membutuhkan bantuan, pertukaran pelajar mandiri menawarkan pilihan perkuliahan hingga 20 SKS, baik yang mencakup mata kuliah Tinggi Penerima Perguruan Tinggi maupun kegiatan dalam bentuk Modul Nusantara. Selain itu, siswa ditawarkan pilihan untuk mendaftar maksimal 6 SKS pendidikan tinggi secara online. Metode yang digunakan dalam penelitian ini adalah algoritma svm, dataset yang digunakan terdiri dari 1000 ulsan komentr dengan perbandingan 70:30. Penelitian ini diimplementasikan dalam sistem web menggunakan bahasa pemrograman Python. Dari 300 data uji yang diimplementasikan dengan 700 data latih. Algoritma Support Vector Machine (SVM) dalam mengklasifikasi data review mendapatkan akurasi tertinggi pada pembagian data latih & data uji 70:30 sebesar 85.00% lalu presisi 28.33%, recall 33.33%.

ABSTRACT

Keywords: Support Vector Machine, student exchange

Support Vector Machine (SVM) is a machine learning method that is widely used for regression and classification problems, especially application review classification. Student exchange is one of the programs that universities must prepare. The student exchange program is intended to reduce the problem of disparities in educational facilities and infrastructure in Indonesia. The advantage of student exchange is that they can manage their time, have high awareness in communicating, are able to admit when they experience problems and need help, independent student exchange offers study options of up to 20 credits, both covering Higher Education Recipients courses and activities in the form of the Nusantara Module. Additionally, students are offered the option to register for a maximum of 6 credits of higher education online. The method used in this research is the SVM algorithm, the dataset used consists of 1000 comment reviews with a ratio of 70:30. This research was implemented in a web system using the Python programming language. Of the 300 test data implemented with 700 training data. The Support Vector Machine (SVM) algorithm in classifying review data obtained the highest accuracy in dividing training data & test data 70:30 at 85.00% then precision 28.33%, recall 33.33%.

1. INTRODUCTION

Pada Analisis sentimen untuk mengumpulkan pendapat orang tentang setiap insiden dunia nyata. Memahami apa yang diperjuangkan individu secara emosional sangat penting di saat-saat seperti yang dialami dunia saat ini (Chakraborty et al., 2020). Analisis sentimen mengumpulkan data ini dan menggunakannya untuk membangun sistem yang dapat mengkategorikan pesan dalam dokumen. analisis sentimen digunakan untuk menilai pendapat tentang suatu peristiwa. Menggunakan penambangan teks, pemrosesan bahasa alami, dan analisis etimologis, analisis sentimen memeriksa pendapat tentang barang atau jasa tertentu. Informasi yang

dihasilkan dari analisis sentimen dibagi menjadi nilai positif dan negatif (Hidayat, 2015).

Program pertukaran pelajar dalam negeri selama 1 (satu) semester yang akan mengundang mahasiswa baru Indonesia untuk merasakan kesempatan belajar terbaik di PT di tanah air. Selain itu, siswa mendapat kesempatan untuk mengenal budaya Indonesia secara langsung melalui tulisan dan praktik. Mengikuti Pertukaran Mahasiswa Merdeka akan memberikan beberapa keuntungan, antara lain kesempatan untuk: (1) menggali dan mempelajari kekayaan budaya nusantara; (2) berteman dengan siswa dari daerah lain; dan (3) belajar di kampus lain di Indonesia. Untuk mengetahui tingkat keberhasilan program (Sartika Pasaribu et al., 2023).

Support Vector Machine adalah teknik klasifikasi yang membagi dua kelompok data pada ruang input dengan menggunakan ide *hyperplane*. Menemukan *hyperplane* optimal dengan margin terbesar antara kedua kelas adalah tujuan dari SVM. SVM dapat memberikan generalisasi yang lebih tinggi dalam klasifikasi dengan memiliki margin terbesar. Dalam SVM, *hyperplane* adalah bidang atau garis yang membagi kelas menjadi dua kelas data. Dengan memanfaatkan teknik pengukuran *edge hyperplane*, SVM mencari *hyperplane* yang paling dapat membedakan kedua kelas. Data terdekat, disebut sebagai *support vector*, ditemukan pada titik terbesar dari *edge hyperplane*. Vektor dukungan digunakan oleh SVM untuk menemukan data yang paling penting (Rosdiana, Ula, & Aidilof, 2021).

Penelitian yang dilakukan oleh Dani Saputra dan Maxsi Ary, 2020, Prediksi minat klient pada produk deposito menggunakan algoritma SVM parameter kernel *polynomial*. Penelitian ini dilakukan untuk mencari prediksi minat klien produk deposito. Penelitian ini menghasilkan data akurasi dari metode *support vector machine* dengan teknik *k-fold cross validation*, mencapai tingkat sekitar 94,91%, dan nilai AUC sebesar 0,979 dengan menggunakan validasi *10-fold*. Selanjutnya, setelah dilakukan perhitungan lebih lanjut dengan menggunakan parameter kernel polinomial, diperoleh tingkat akurasi yang lebih tinggi, yakni sebesar 95,18%, dengan nilai AUC mencapai 0,974 (Saputra & Ary, 2020).

Dibandingkan dengan algoritma klasifikasi lainnya, SVM memiliki gagasan matematika yang lebih berkembang dan mudah dipahami. SVM cocok untuk digunakan dalam konsep analisis karena dapat menyelesaikan masalah klasifikasi dan regresi menggunakan model linier dan non-linier. Berdasarkan latar belakang diatas, penulis melakukan penelitian yang berjudul “ **Analisis Sentimen Terhadap Program Pertukaran Mahasiswa Kampus Merdeka menggunakan Metode Support Vector Machine (SVM)**” Penelitian ini nantinya akan memberi gambaran umum mengenai pandangan masyarakat mengenai Pertukaran Mahasiswa Kampus Merdeka cenderung mengarah ke positif, negatif, netral, berdasarkan analisis sentiment terhadap data yang di ambil melalui *instagram* menggunakan metode *Support Vector Mechine (SVM)*.

2. RESEARCH METHODS

1. Pengumpulan data

Pengambilan data menggunakan secara langsung pada situs *instagram* melalui akun kampus Merdeka dengan *hashtag*

		Prediksi		
		Positif	Negatif	Netral
Actual	Positif	True Positif	False Negatif	False Netral
	Negatif	False Positif	True Negatif	False Netral
	Netral	False Positif	False Negatif	True Netral

pertukaran mahasiswa dengan mengambil komentar komentar pada postingan secara langsung.

2. Labelisasi

Berdasarkan rating sebagai data latih, dilakukan pelabelan untuk memberikan ID secara manual pada setiap bagian data komentar. Kategori kelas positif dan negatif disertakan di sini. Keluaran pelabelan adalah label.data di *csv*.

3. Pengolahan Data

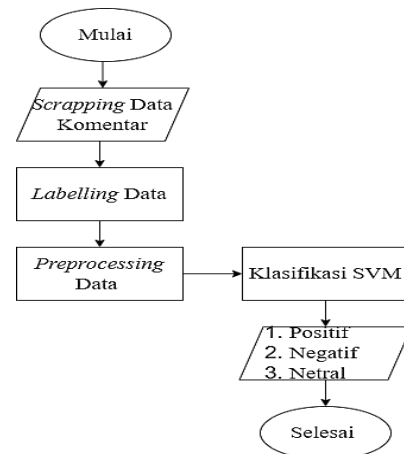
Langkah persiapan pemrosesan data adalah ketika kata dan teks yang ada dalam data dievaluasi untuk format dan karakter non-huruf dihilangkan. Membersihkan data, menyalarkan teks menjadi huruf kecil, normalisasi, memecah teks menjadi token, menghapus kata-kata yang berlebihan (*stopword removal*), dan *stemming* adalah beberapa prosedur yang dilakukan selama tahap pemrosesan.

Data kemudian harus diberi bobot untuk membuat angka saat tahap *preprocessing* selesai. Pembobotan data dilakukan agar data lebih mudah untuk diinterpretasikan oleh sistem klasifikasi. Setelah itu, semua data yang tersedia akan dipecah menjadi data pelatihan dan data uji dengan rasio 70:30. Algoritma klasifikasi SVM akan menganalisis data uji dan memberikan prediksi.

4. Uji Coba Sistem dengan Manual

Uji coba merupakan tahapan dalam melakukan identifikasi hasil dari klasifikasi sentimen yang telah dilakukan. Pada bagian ini akan dilakukan pembahasan terkait hasil klasifikasi dari sistem yang telah dibangun dengan hasil perhitungan yang dilakukan secara manual

2.1 Skema Sistem



1. Mengumpulkan data komentar dari *instagram* menggunakan *hashtag* pertukaran mahasiswa pada postingan
2. Beri label data, pada tahap ini pemberian label dilakukan secara manual.
3. Melakukan *preprocessing* pada data yang sudah diberi label.
4. Melakukan klasifikasi SVM
5. Menampilkan hasil akhir klasifikasi, *Output* yang dihasilkan dari klasifikasi yaitu data komentar yang termasuk kedalam kelas positif, negatif dan netral.

1. confusion matriks

Confusion matrix merupakan alat yang digunakan untuk mengukur tingkat ketepatan suatu klasifikasi dalam konsep *data mining*. Akurasi diukur sebagai persentase dari data latih yang berhasil diklasifikasikan dengan benar. Hasil pengujian dari suatu metode dianalisis melalui tabel (*confusion matrix*) untuk mendapatkan nilai-nilai seperti akurasi, presisi, dan *recall*. *Confusion matrix* memberikan gambaran rinci tentang kesalahan prediksi yang terjadi.

Aktual merupakan nilai asli dari *label class*. Sedangkan prediksi merupakan nilai prediksi hasil pemodelan *machine learning*. Berdasarkan tabel di atas:

a. *True* positif (TP) merupakan jumlah *record* data positif yang diklasifikasikan sebagai nilai positif.

b. False Positif (FP) yaitu jumlah *record* data negatif yang diklasifikasikan sebagai nilai positif.

c. False Positif (FP_t), yaitu jumlah *record* data netral yang diklasifikasi sebagai nilai positif.

d. False negatif (FN) yaitu jumlah *record* data positif yang diklasifikasikan sebagai nilai negatif.

e. True Negatif (TN) yaitu jumlah *record* data negatif yang diklasifikasikan sebagai nilai negatif.

f. False Negatif (FN_t) yaitu jumlah *record* data netral yang diklasifikasikan sebagai nilai negatif.

g. False Netral (FN_t) yaitu jumlah *record* data positif yang diklasifikasikan sebagai nilai netral.

h. False Netral (FN_t) yaitu jumlah *record* data negatif yang diklasifikasikan sebagai nilai netral.

i. True Netral (TN_t) yaitu jumlah *record* data netral yang diklasifikasikan sebagai nilai netral.

3. RESULT AND DISCUSSION

Analisis sistem

Sistem yang akan diimplementasikan merupakan sebuah sistem untuk membantu menganalisis komentar-komentar pada pertukaran mahasiswa kampus merdeka pada akun *instagram*. Sistem analisis sentimen ini dibangun berbasis web dengan menggunakan bahasa python dan menggunakan framework flask untuk tampilannya. Cara kerja sistem ini melalui beberapa tahapan yaitu pengumpulan data yang diambil dari kolom komentar *instagram* menggunakan teknik instant data. Kemudian diberi label secara manual yaitu positif, negatif dan netral.

Selanjutnya data yang sudah diberi label akan dilakukan *preprocessing* untuk menghilangkan kata-kata yang tidak diperlukan. Kemudian data dibagi menjadi data latih dan data uji. Selanjutnya akan dilakukan pembobotan kata dengan TF-IDF yang hasil perhitungannya akan dimasukkan ke dalam klasifikasi dengan algoritma *Support Vector Machine*. Hasil klasifikasi data uji kemudian akan menghasilkan data yang dikategorikan ke dalam kelas positif, dan kelas negatif.

Analisis data

Berdasarkan kebutuhan yang digunakan pada sistem ini, data diperoleh dari review atau komentar pada pertukaran mahasiswa kampus merdeka pada akun *instagram*. Proses pengumpulan data dilakukan dengan menggunakan teknik teknik instant data

Data hasil *scraping* keseluruhannya berjumlah 1000 data, data tersebut dijadikan dataset dengan format file dataset yang digunakan adalah .csv.

Kemudian dilakukan pemberian label secara manual berdasarkan rating yang diberikan para pengguna layanan aplikasi tersebut ke dalam kelas positif, negatif, dan netral.

No	Term	Tf		
		K1	K2	K3
1	Besok	0,477	0	0
2	Itu	0,477	0	0
3	Tutup	0,477	0	0
4	Langsung	0,176	0	0
5	terus	0,477	0	0

6	ada	0,477	0	0
7	umum	0,477	0	0
8	lolos	0,477	0	0
9	berkas	0,477	0	0
10	tahap	0,477	0	0
11	min	0,477	0	0
12	kayaknya	0	0,477	0
13	full	0	0,477	0
14	mata kuliah	0	0,477	0
15	tidak	0	0,477	0
16	buka	0	0,477	0
17	leptop	0	0,477	0
18	program	0	0	0,477
19	elit	0	0	0,477
20	nuntasin	0	0	0,477
21	benefit	0	0	0,477
22	pmm	0	0	0,477
23	sulit	0	0	0,477

Evaluasi klasifikasi sistem

Data training yang digunakan yaitu 700 komentar, dan data testing yang digunakan sebanyak 300 komentar. Berdasarkan data testing tersebut nilai aktual positif yang diuji sebanyak 255, aktual negatif sebanyak 6, aktual netral sebanyak 39.

Aktual		Positif	Negatif	Netral
	Positif	255	0	0
	Negatif	6	0	0
	Netral	39	0	0

Aktual merupakan nilai asli dari label atau class, sedangkan prediksi hasil pemodelan machine learning dari tabel tersebut dapat disimpulkan bahwa :

1. Dari nilai aktual positif sebanyak 255 data menghasilkan :

$$\text{Akurasi} = \frac{255 + 0 + 0}{255 + 6 + 39} \times 100\% \cdot \frac{255}{300} \times 100\% = 85\%$$

Selanjutnya mencari nilai presisi pada setiap kelas sebagai berikut :

$$\text{Presisi (Positif)} = \frac{255}{255 + 6 + 39} = \frac{255}{300} = 0.86$$

$$\text{Presisi (Negatif)} = \frac{TN}{TN, FP, N_t N_g} = \frac{0}{0+0+0} = \frac{0}{0} = 0$$

$$\text{Presisi (Netral)} = \frac{TN_t}{TN_t + PN_t + N_g N_t} \cdot \frac{0}{0+0+0} = \frac{0}{0} = 0$$

$$\text{Presisi} = \frac{\text{Presisi (POs)} + \text{Presisi (neg)} + \text{Presisi (Net)}}{3} \times 100\% = \frac{0,85 + 0 + 0}{3} \times 100\% = 28,3$$

Perhitungan recall tiap tiap kelas sebagai berikut

$$\begin{aligned} \text{recall(Positif)} &= \frac{T_p}{TP+PN_t+FP} = \frac{255}{255} = 1 \\ &= 0.86 \\ \text{recall (Negatif)} &= \frac{TN_t}{TN_g+FN+N_tN_g} = \frac{0}{0+0+0} = \frac{0}{0} = 0 \\ \text{recall (Netral)} &= \frac{TN_t}{TN_t+N_t+N_gN_t} = \frac{0}{0+0+0} = \frac{0}{0} = 0 \end{aligned}$$

Maka hasil recall perhitungan keseluruhannya sebagai berikut :

$$\begin{aligned} \text{recall} &= \frac{\text{Recall (POs)} + \text{Recall (neg)} + \text{Recall (Net)}}{3} \times 100\% \\ &= \frac{1 + 0 + 0}{3} \times 100\% = 33.33\% \end{aligned}$$

$$\text{recall (Negatif)} = \frac{TN_t}{TN_g+FN+N_tN_g} = \frac{0}{0+0+0} = \frac{0}{0} = 0$$

$$\text{recall (Netral)} = \frac{TN_t}{TN_t+N_t+N_gN_t} = \frac{0}{0+0+0} = \frac{0}{0} = 0$$

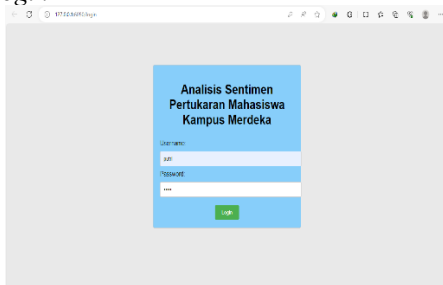
Maka hasil recall perhitungan keseluruhannya sebagai berikut :

$$\begin{aligned} \text{recall} &= \frac{\text{Recall (POs)} + \text{Recall (neg)} + \text{Recall (Net)}}{3} \times 100\% \\ &= \frac{1 + 0 + 0}{3} \times 100\% = 33.33\% \end{aligned}$$

Langkah penerapan analisis sentimen melibatkan implementasi sistem dengan menggunakan bahasa pemrograman berdasarkan hasil analisis yang telah dilakukan. Dalam sistem ini, terdapat beberapa halaman seperti halaman login, dashboard, preprocessing, split data, klasifikasi SVM, score performance, dan Wordcloud.

1. Halaman Login

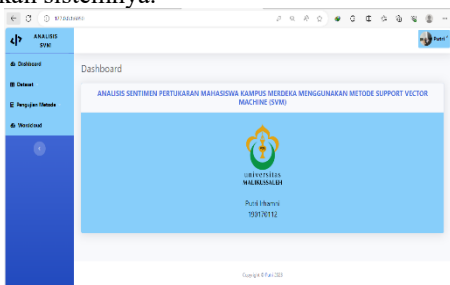
Halaman *login* merupakan halaman yang berisi form untuk admin melakukan *login*. Terdapat masukkan username dan password yang harus admin isi untuk bisa melakukan *login*. Jika username dan password salah, maka akan muncul notifikasi *login* gagal, sehingga admin akan Kembali ke 1. halaman *login*



Gambar 1 halaman login

2. Halaman Dashboard

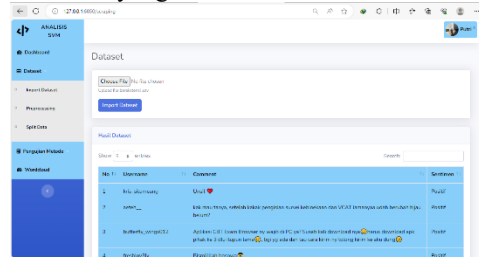
Halaman dashboard merupakan halaman utama yang mempunyai beberapa menu untuk memudahkan admin menjalankan sistemnya.



Gambar 2 Halaman Dashboard

3. Halaman Import Dataset

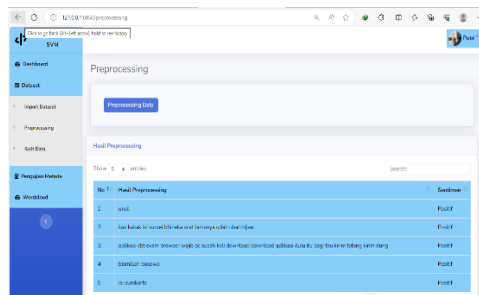
Halaman import dataset merupakan halaman untuk memasukkan data yang akan di analisis sentimen.



Gambar 3. Halaman Import Dataset

4. Halaman Preprocessing

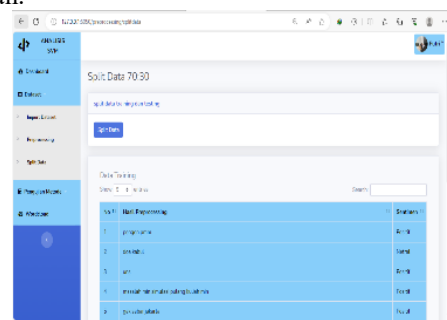
Halaman ini merupakan halaman untuk melakukan *preprocessing*.



Gambar 4 Halaman Preprocessing

5. Halaman Split Data

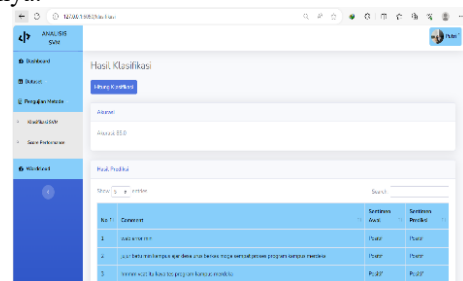
Halaman split data merupakan halaman untuk menampilkan data *training* dan data testing yang telah dipisahkan.



Gambar 5 Halaman Split Data

6. Halaman Klasifikasi SVM

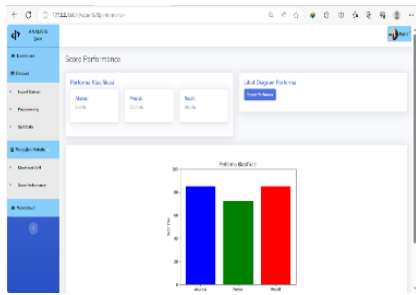
Halaman ini merupakan halaman untuk menampilkan data klasifikasi *Support Vector Machine* yang telah di analisis sentimen oleh sistem, dan dibandingkan antara actual dengan prediksinya.



Gambar 6 Halaman Klasifikasi SVM

7. Halaman Score Performance

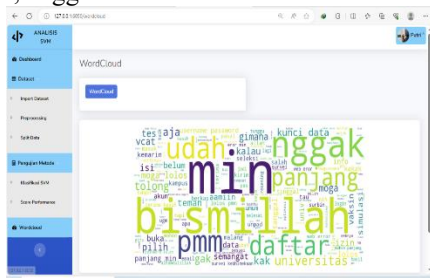
Halaman ini merupakan halaman untuk menampilkan score performance. Adapun nilai performanya yaitu akurasi 85%, precision 72,25%, recall 85%.



Gambar 7 Halaman Score Performance

8. Halaman Wordcloud

Halaman ini merupakan halaman untuk menampilkan Wordcloud. Adapun kata (Wordcloud) urutan pertama, kedua dan ketiga yang paling banyak muncul yaitu “min”, “bismillah”, “nggak”.



Gambar 8 Halaman Wordcloud

4. CONCLUSION

Dari 1000 sampel komentar pada pertukaran mahasiswa kampus Merdeka pada akun *Instagram* memiliki nilai positif 857 ulasan dan kelas negatif 21 ulasan, dan netral 122 ulasan, Hasil penerapan *Support Vector Machine* pada klasifikasi data *training* ulasan pada pertukaran mahasiswa kampus Merdeka pada akun *Instagram* mendapatkan akurasi tertinggi pada pembagian data latih dan data uji 70:30 sebesar 85%, presisi 28.3%, dan recall 33,33%, penelitian analisis sentimen mengenai komentar pada pertukaran mahasiswa kampus merdeka pada akun *instagram* menggunakan metode *Support Vector Machine* (SVM) ini melibatkan dataset sebanyak 1000 data yang di implementasikan dalam sistem web.

REFERENCES

[1] Arifin, Z. (2019). Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql. *NJCA (Nusantara Journal of Computers and Its Applications)*, 4(1). <https://doi.org/10.36564/njca.v4i1.101>

[2] Bari, M., Sitorus, S. H., Ristian, U., Rekayasa, J., Komputer, S., Mipa, F., Tanjungpura, U., Prof, J., Hadari, H., & Pontianak, N. (2018). Penyebaran Wabah Penyakit ISPA (Studi Kasus: Wilayah Kota Pontianak).

[3] Abdurrazzaq, M. A. (2023). Analisis Ulasan Aplikasi MyPertamina Menggunakan Topic Modeling dengan Latent Dirichlet Allocation. 10(1), 1–6.

[4] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers— A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing Journal*, 97, 106754. <https://doi.org/10.1016/j.asoc.2020.106754>

[5] Darnila, E., Al-Kautsar, H., & Iksan, Y. (2020). Analysis of Patient Disease Trends Based on Medical Record Data Using the C4. 5 Algorithm. *Login: Jurnal Teknologi*

[6] Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>

[7] Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66.

[8] Hidayat, andi nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik)*, 1(1), 1–7.

[9] Khaidarni, U., Fauzi, A. H., Nisa, H., Gumelar, R., Muldi, A., & Sultan, U. (2023). Analisis Difusi Inovasi Terhadap Aplikasi Mypertamina. *Jurnal Ilmiah Wahana Pendidikan*, Januari, 2023(2), 644–652.

[10] Maria, R., Umayah, R. U., Mahardinny, S., Kalana, D. N., & Saputra, D. D. (2023). Analisis Sentimen Persepsi Masyarakat Terhadap Penggunaan Aplikasi My Pertamina Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier. 1, 1–10.

[11] Munirul, Ula, Alvanof, M. M., & Triandi, R. (2020). Analisa Dan Deteksi Konten Hoax Pada Media Berita. *Jurnal Teknologi Terapan & Sains 4.0 Universitas Malikussaleh*, 1, 2.

[12] Najib, A. C., Irsyad, A., Qandi, G. A., & Rakhmawati, N. A. (2019). Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. *Fountain of Informatics Journal*, 4(2), 41. <https://doi.org/10.21111/fij.v4i2.3573>

[13] Purnomo, L. M., & Ayub, M. (2021). Analisis data hasil web scraping untuk menentukan kualitas jurnal ilmiah. *Jurnal STRATEGI-Jurnal Maranatha*, 3(1), 122–132.

[14] Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(2), 802–810.

[15] Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/10.22146/jnteti.v9i2.102>

[16] Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Health Information management Journal*, 8(2), 100–109.

[17] Retno, S., Dinata, R.K., Hasdyna, N. (2023). Evaluasi model data chatbot dalam natural language processing menggunakan k-nearest neighbor. *Jurnal CoSciTech (Computer Science and Information Technology)*. 4(1): 146-153.

[18] Rosmala, D., & Dwipa, G. (2012). Pembangunan Website Content Monitoring System Menggunakan System Menggunakan DiffliB Python. *Informatika*, 3(3), 1–6. www.Telkom.com

[19] Saputra, D., & Ary, M. (2020). Prediksi Minat Klien pada Produk Deposito Menggunakan Algoritma SVM Parameter Kernel Polynomial. *EProsiding Sistem Informasi (POTENSI)*, 1(1), 162–171.

[20] Yulita, W. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 1. <https://doi.org/10.33365/jdmsi.v2i2.1344>.