Jurnal Serina Sains, Teknik dan Kedokteran
Vol. 01, No. 02, Sep 2023 : hlm 237-246 ISSN-L 2986-5204 (Versi Elektronik)

APLIKASI DETEKSI KOMENTAR SPAM YOUTUBE DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS WEB

Raymond Tjahyadi[1], Viny Christanti Mawardi[2], Novario Jaya Perdana[3]

[1]Jurusan Teknik Informatika, Universitas Tarumanagara Jakarta
Email: raymond.535190030@stu.untar.ac.id
[2]Jurusan Teknik Informatika, Universitas Tarumanagara Jakarta
Email: vinym@fti.untar.ac.id
[3]Jurusan Teknik Informatika, Universitas Tarumanagara Jakarta
Email: novariojp@fti.untar.ac.id

Masuk : 24-11-2022, revisi: 12-12-2022, diterima untuk diterbitkan : 19-12-2022

ABSTRAK

YouTube merupakan salah satu platform media sosial terbesar di dunia dengan jumlah pengguna 2,6 milyar di seluruh dunia dan 192 juta pengguna di seluruh indonesia. Hal ini menciptakan peluang bagi kejahatan di dunia digital. Spam adalah salah satu tindakan yang dapat merugikan pengguna Youtube dan pihak yang juga mengalami kerugian adalah pemilik channel karena banyak komentar spam yang mengganggu sehingga dapat membuat kenyamanan para penonton berkurang. Selain daripada pemilik channel yang dirugikan spam juga merugikan pihak penonton karena komentar spam dapat memberikan informasi yang salah, menimbulkan distraksi, mencuri informasi dan memancing emosi sehingga menyebabkan pertengkaran. Klasifikasi merupakan metode yang dapat membedakan jenis komentar spam dan komentar non-spam. Metode *support vector machine* adalah salah satu metode klasifikasi yang mampu membedakan jenis kategori dengan baik dan dapat menghasilkan prediksi yang akurat serta tingkat akurasi yang tinggi. Terdapat beberapa kernel yang dapat melakukan klasifikasi yaitu linear, polynomial, RBF, dan sigmoid. Adapun pendekatan *one vs one* dan *one vs rest* untuk klasifikasi *multiclass*. Dilakukan 5 kali pengujian terhadap model *one vs one*, *one vs rest,* RBF*, dan polynomial dengan menggunakan 581 data yang telah dikumpulkan* untuk mencari model yang terbaik. Model terbaik dari hasil pengujian adalah model algoritma SVM dengan pendekatan *one vs one dengan akurasi* sebesar 89,58% dengan pemisahan 75% data latih dan 25% data uji.

Kata kunci: Akurasi, klasifikasi, model, prediksi, spam

ABSTRACT

YouTube is one of the largest social media platforms in the world with 2.6 billion users worldwide and 192 million users throughout Indonesia. This creates opportunities for crime in the digital world. Spam is one of the actions that can harm Youtube users and those who also suffer losses are channel owners because there are many spam comments that are annoying so that they can reduce the comfort of the viewers. Apart from the channel owners who are harmed by spam, it also harms the audience because spam comments can provide wrong information, cause distraction, steal information and provoke emotions, causing fights. Classification is a method that can distinguish the types of spam comments and non-spam comments. The Support Vector Machine method is a classification method that is able to distinguish between types of categories well and can produce accurate predictions and a high degree of accuracy. There are several kernels that can classify, such as linear, polynomial, RBF, and sigmoid. The One vs One and One vs Rest approaches for multiclass classification. Five tests were carried out on the One vs One, One Vs Rest, RBF, and Polynomial models using 581 collected data to find the best model. The best model from the test results is the SVM algorithm model with the One vs One approach with an accuracy of 89.57% with a separation of 75% training data and 25% test data.

Keywords: *Accuracy, classification, model, predictions, spam*

1. PENDAHULUAN

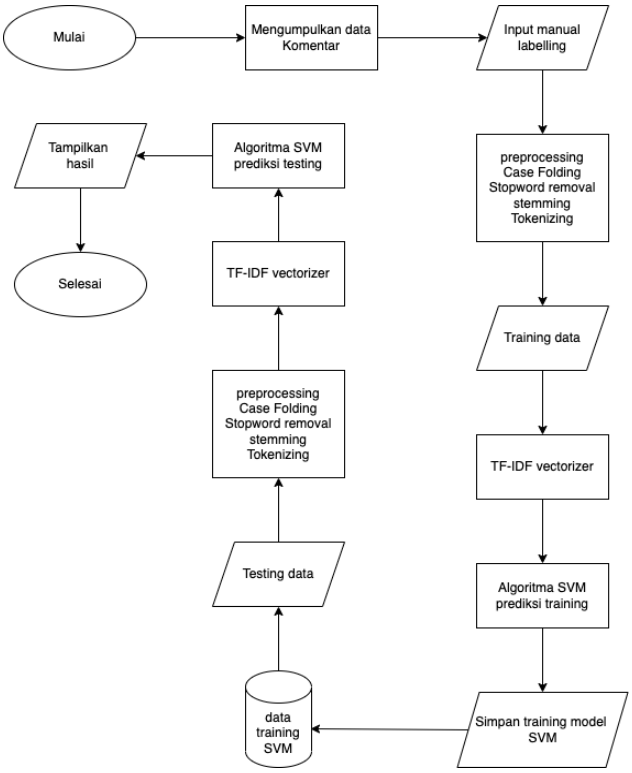
Spam didefinisikan konten-konten yang tidak sesuai atau mengandung topik yang tidak relevan dengan topik yang ada (Julio, R. 2022). Spam juga diartikan sebuah konten yang tidak diperlukan namun secara berulang kali muncul dan memberikan efek negatif pada pengguna secara umum (Aji, F. P., 2018). Beberapa orang yang melakukan tindakan spam (*spammer)* menggunakan komentar youtube dengan tujuan untuk mempromosikan *channel*, tindakan penipuan, dan mengatasnamakan pembuat channel menggunakan *spam bot* untuk mengirim pesan otomatis secara massal. Komentar spam yang sering muncul berulang kali di kolom komentar jika tidak terdeteksi atau disaring dengan benar, *channel* tertentu akan kehilangan popularitas dan pengikut *channel* tersebut akan menurun karena banyak spam yang mengganggu. Hal ini dimanfaatkan oleh orang-orang yang ingin mendapatkan keuntungan dengan cara yang ilegal yaitu, komentar spam.

Penelitian terdahulu telah melakukan klasifikasi komentar spam pada Instagram menggunakan metode *support vector machine* dan mendapatkan nilai akurasi sebesar 97.90% (Syam, A. T., 2020). Kelebihan dari *support vector machine* adalah kemampuan untuk menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit (Ichwan, M., 2018). Penelitian lainnya telah melakukan implementasi algoritma *multiclass* SVM pada opini publik bahasa Indonesia di Twitter mendapatkan hasil kesimpulan Model One vs One lebih unggul 0,06 dibandingkan model one vs rest (Alita, D., 2020). Text Mining adalah proses penemuan pengetahuan menggunakan *natural language processing* (NLP) dengan cara menggali informasi dari sebuah data berformat teks (Rosari, M. A., 2022). *Text mining* mampu menghasilkan informasi melalui tahap pemrosesan, pengelompokan, dan analisis dari data-data yang tidak terstruktur dalam jumlah besar. Teks *preprocessing* adalah tahapan yang digunakan untuk mengubah data teks menjadi bentuk yang lebih mudah dipahami. Tahap *preprocessing* terdiri dari *case folding*, tokenisasi, menghapus *stopwords*, dan *stemming*. *Case folding* adalah sebuah proses mengubah huruf besar ke huruf kecil. Tokenisasi adalah proses mengubah data teks menjadi token pemrosesan terhadap data teks untuk diubah kedalam bentuk token atau kata yang sering disebut dengan tokenisasi. Data dalam bentuk dokumen, paragraf ataupun kalimat dipecah menjadi token-token sebagai masukan untuk proses selanjutnya yaitu *text transformation*. Menghapus *stopwords* adalah sebuah proses penghilangan pada kata sambung yang sangat umum berada pada dokumen. Kata sambung tersebut dihilangkan karena terlalu bersifat umum sehingga tidak memiliki dampak yang signifikan terhadap dokumen. Proses *stemming* adalah menghapus imbuhan pada token sehingga hasil yang didapat merupakan kata dasar dari kata tersebut. Proses *stemming* dilakukan untuk mengurangi jumlah daftar kata pada indeks. Setelah tahap preprocessing dilakukan *term frequency-inverse document frequency* (TF-IDF) untuk menghitung bobot setiap kata yang paling umum digunakan. *Term frequency-inverse document frequency* (TF-IDF) adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen atau dalam sekelompok kata (Thomas, V. 2022).

Dari permasalahan yang dijelaskan pada bagian latar belakang penulis mendapatkan beberapa rumusan masalah yaitu: (a) apakah definisi dari spam?; (b) bagaimana cara membedakan komentar spam?; (c) apa kelebihan dan kelemahan algoritma *support vector machine*?; (d) bagaimana melakukan klasifikasi komentar Spam?.

https://doi.org/10.24912/jsstk.v1i2.28510

2. METODE PENELITIAN



Gambar 1. Flowchart alur pembangunan sistem deteksi spam

Flowchart dari gambar diatas merupakan tahapan pembangunan sistem deteksi komentar spam menggunakan algoritma SVM sebagai berikut:

(a) Pengumpulan Data
Data dikumpulkan menggunakan kode python yang ditulis menggunakan *visual studio code* yang menggunakan Youtube API v3 untuk mengumpulkan 581 data dari 4 youtubers *channel yaitu,* Jerome polin, jess no limit, Gadgetin, dan tanboy. Jerome polin merupakan *channel* yang membahas tentang edukasi dan hiburan, jess no limit merupakan *channel* yang membahas tentang gaming dan hiburan, gadgetin merupakan *channel* yang membahas tentang gadget dan tanboy merupakan *channel* yang membahas tentang makanan. Terdapat banyak pilihan data yang dapat digunakan untuk menjadi data latih dalam penelitian ini akan tetapi data yang dipilih adalah dari 4 *channel* tersebut karena *channel* tersebut memiliki penggemar yang banyak dan karya mereka memiliki pengaruh yang besar bagi masyarakat Indonesia.

(b) Labelling Data
Tahapan labelling data untuk memberikan label data promosi, tautan dan bukan spam pada data yang telah dikumpulkan dari proses pengumpulan data. Tujuan dari proses labelling data adalah supaya dapat melatih model SVM dalam klasifikasi kategori komentar spam dan bukan spam. Proses labelling data ini dilakukan secara manual dengan menggunakan alat bantu microsoft excel.

(c) Pembersihan Data
Tahap pembersihan data terdiri dari proses *case folding* yaitu mengubah keseluruhan kalimat menjadi huruf kecil, data *cleaning* untuk mencari *missing value* pada dataset, penghapusan simbol, *stopwords*, tautan, spasi kosong dan *emoticon.* Setelah itu melakukan tokenisasi, *stemming* dan

pembobotan TF-IDF agar data dapat siap diuji pada tahap berikutnya. Proses pembersihan data akan dilakukan dengan menggunakan *jupyter notebook* setelah data sudah diberikan label.

(d) Pengujian Data

Pengujian data merupakan tahap sesudah data dibersihkan dan sudah dilakukan vektorisasi TF-IDF. Tahap ini dibagi menjadi 2 tahap yaitu, uji *data training* dan uji *data testing*. Dalam tahap uji *data training* akan diuji menggunakan metode *support vector machine* dan *confusion matrix* untuk menampilkan evaluasi dari model *support vector machine*. Proses pengujian data *training* akan dilakukan dengan menggunakan *jupyter notebook*. Dalam tahap uji *data testing* akan menggunakan data komentar dari URL yang dimasukan dalam aplikasi. Uji data testing dilakukan pada *website* yang akan dibangun.

Support vector machine (SVM) merupakan salah satu model dalam algoritma *machine learning*. SVM bekerja dengan mendefinisikan hyperplane (fungsi pembatas yang dapat digunakan untuk pemisah antar kelas) yang memaksimalkan margin antara dua kelas yang berbeda. Garis pemisah terbaik antara dua kelas yang berbeda atau disebut juga *hyperplane* dapat dicari nilainya dengan mencari nilai terdekat antara kelas -1 dengan *hyperplane* dan juga nilai terdekat antara +1 dengan hyperplane (Jihad, 2020).

Rumus Persamaan *Hyperplane*:

$$f(x) = \vec{w} \,.\, x + b \qquad (1)$$

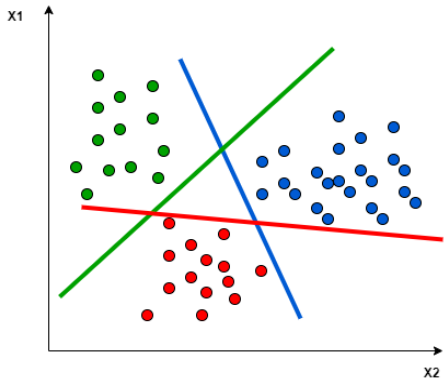
Keterangan:

w : suatu bobot vektor, yaitu $\{w1, w2, \ldots, wn\}$
n : jumlah atribut
b : suatu skalar yang disebut dengan bias.

SVM juga memiliki beberapa kekurangan dalam mengklasifikasi lebih dari 2 kategori karena SVM awalnya dikenal sebagai metode yang baik untuk klasifikasi biner. Oleh karena permasalahan Spam yang diteliti pada penelitian ini memiliki 3 kategori maka akan digunakan pendekatan *multiclass* SVM.

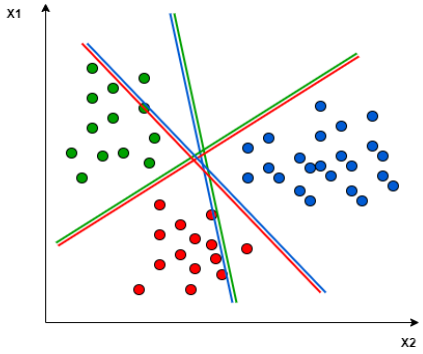
Multiclass support vector machine melakukan klasifikasi menggunakan prinsip menyederhanakan permasalahan multiclass menjadi beberapa permasalahan *binary*. Prinsip multiclass SVM terbagi menjadi 2 prinsip yaitu: (1) *one vs rest.* Dalam pendekatan *one vs rest* dibutuhkan sebuah hyperplane yang memisahkan sebuah kelas dengan semua kelas yang lainnya. Sebagai contoh sederhana garis hijau memaksimalkan pemisahan antara titik hijau dengan kelas biru dan merah sekaligus. Garis merah memaksimalkan antara titik merah dengan kelas hijau dan biru berlaku juga untuk garis biru. Gambar ilustrasi *one vs rest* dapat dilihat pada gambar 2.

https://doi.org/10.24912/jsstk.v1i2.28510



Gambar 2. one vs rest
Sumber Gambar: Baeldung, (2022, September 25)

(2) *one vs one*. Dalam pendekatan *one vs one* dibutuhkan hyperplane untuk memisahkan antara setiap dua kelas dengan mengabaikan titik dari poin ketiga. Sebagai contoh sederhana garis merah-biru mencoba memaksimalkan untuk pemisah hanya merah-biru tidak ada hubungannya dengan data warna hijau. garis biru-hijau mencoba memaksimalkan untuk memisahkan biru dan hijau tidak ada hubungannya dengan data merah. Gambar ilustrasi *one vs one* dapat di lihat pada gambar di bawah:



Gambar 3. one vs one
Sumber Gambar: Baeldung, (2022, September 25)

Pada tahap evaluasi, dilakukan pengujian metriks akurasi, *recall*, *precision*, dan F-1 *score* terhadap model yaitu akurasi atau *accuracy* adalah tolak ukur yang mengukur seberapa akurat model dalam memprediksi dengan benar (Iqbal, A. R., 2022). Akurasi dapat dirumuskan sebagai berikut:

$$accuracy = \frac{(TN + TP)}{(TN + TP + FP + FN)} \qquad (2)$$

Keterangan:
TN: *True Negative*
TP: *True Positive*
FN: *False Negative*
FP: *False Positive*

(1) *Recall*
Recall adalah Perbandingan antara *true positive* dengan banyaknya data yang bernilai positif. Semakin kecil *false negative (FN)* membuat recall semakin besar. *Recall* dapat dirumuskan sebagai berikut:

$$Recall = \frac{TP}{(TP + FN)} \qquad (3)$$

(2) *Precision*
Precision adalah perbandingan antara *true positive* dengan banyaknya data yang diprediksi positif. semakin kecil *false positive* (FP), membuat *precision* semakin besar. *Precision* dapat dirumuskan sebagai berikut:

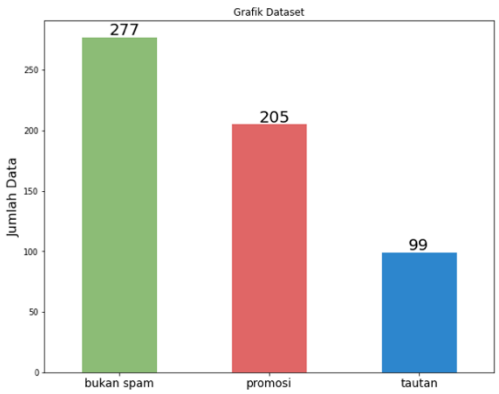
$$precision = \frac{TP}{(TP + FP)} \qquad (4)$$

(3) F-*Measure*
F-Measure atau F-1 score adalah nilai rata-rata antara *recall* dan *precision*. F-*measure* mempunyai nilai terbaik 1.0 dan nilai terburuknya adalah 0. F-1 *score* yang merepresentasikan hasil skor yang baik akan menjadi faktor indikasi bahwa model klasifikasi punya *precision* dan *recall* yang baik. F-*measure* dapat dirumuskan sebagai berikut:

$$F - Measure = \frac{2 * TP}{(2 * TP + FN + FP)} \qquad (5)$$

3. HASIL DAN PEMBAHASAN

Proses pembuatan *website* sistem deteksi spam menggunakan bahasa pemrograman python versi 3.10.6 dan streamlit versi 1.15.0. Penulis melakukan ekstraksi sekitar 581 komentar dari berbagai *channel* di Youtube menggunakan API Youtube dan menyimpannya dalam sebuah file CSV untuk analisis selanjutnya. Khususnya data komentar video dari *channel* Jess No limit, Gadgetin, Jerome Polin, dan Tanboykun yang sedang tren dengan jumlah penayangan yang sangat besar menjadi sasaran. Sampel dataset terdiri dari komentar dari video paling populer dari channel mereka. Penulis secara manual memberi label pada komentar (yaitu memberikan nilai 0/1/2) yang bersifat promosi atau di luar konteks dengan video yang diberikan dan mengklasifikasikannya sebagai spam serta tautan. Berikut merupakan analisis visualisasi dataset dari 581 data yang telah dikumpulkan:



Gambar 4. Dataset spam

https://doi.org/10.24912/jsstk.v1i2.28510

Dilakukan pelatihan 5 kali percobaan untuk menemukan model yang terbaik dengan beberapa pemisahan data latih dengan data uji. Hasil pengujian setiap model dapat dilihat pada tabel berikut:

Tabel 1. Hasil pengujian model one vs rest

Model	Data Latih	Data Uji	Akurasi	Recall	Precision	F1-Score
One vs Rest	70%	30%	85.55%	86.25%	86%	85.13%
One vs Rest	75%	25%	89.58%	91.05%	93.23%	91.62%
One vs Rest	80%	20%	88.7%	89.37%	88.76%	88.51%
One vs Rest	85%	15%	88.51%	89.22%	89.71%	88.11%
One vs Rest	90%	10%	84.48%	87.07%	86.75%	84.04%

Tabel 2. Hasil pengujian model one vs one

Model	Data Latih	Data Uji	Akurasi	Recall	Precision	F1-Score
One vs One	70%	30%	85.55%	86.25%	86%	85.13%
One vs One	75%	25%	89.58%	91.05%	93.23%	91.62%
One vs One	80%	20%	88.7%	89.37%	88.76%	88.51%
One vs One	85%	15%	88.51%	89.22%	89.71%	88.11%
One vs One	90%	10%	84.48%	87.07%	86.75%	84.04%

Tabel 3. Hasil pengujian model kernel polynomial

Model	Data Latih	Data Uji	Akurasi	Recall	Precision	F1-Score
Polynomial	70%	30%	86.13%	86.85%	86.51%	85.77%
Polynomial	75%	25%	88.19%	88.49%	88.52%	87.89%
Polynomial	80%	20%	87.83%	88.79%	87.79%	87.67%
Polynomial	85%	15%	88.51%	89.22%	89.71%	88.11%
Polynomial	90%	10%	84.48%	87.07%	86.75%	84.04%

Tabel 4. Hasil pengujian model RBF

Model	Data Latih	Data Uji	Akurasi	Recall	Precision	F1-Score
RBF	70%	30%	86.71%	87.06%	87.56%	86.23%
RBF	75%	25%	87.5%	87.75%	87.91%	87.13%
RBF	80%	20%	88.7%	89.72%	88.64%	88.59%
RBF	85%	15%	89.66%	90%	91.45%	89.23%
RBF	90%	10%	86.21%	88.41%	89.55%	85.71%

Dari hasil 5 kali pengujian yang telah dilakukan terhadap 4 model, didapatkan model yang terbaik oleh *one vs one* dan *one vs rest* dengan pembagian 75% data latih dan 25% data uji dengan akurasi 89.58% dan nilai F1-Score 91.62%. Model ini kemudian akan diimplementasikan kedalam aplikasi. Gambar di bawah merupakan tampilan aplikasi deteksi komentar spam youtube berbasis *website*:

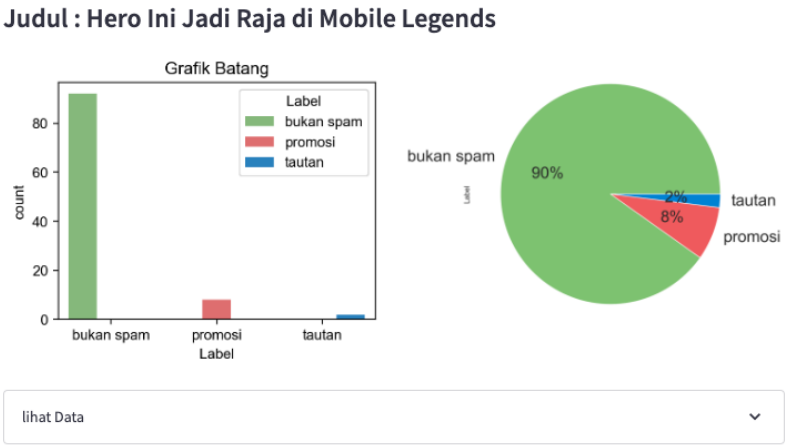
Gambar 5. Tampilan website sistem deteksi spam

Tampilan *website* dibangun menggunakan streamlit sebagai *web framework*. *Streamlit* merupakan *library* untuk membangun website yang mendukung bahasa pemrograman python. Berikut beberapa *libraries* yang digunakan pada tabel di bawah:

Tabel 5. Libraries yang diperlukan

numpy	Selenium	scikit-learn
pandas	Matplotlib	beautifulSoup
streamlit	Sastrawi	NLTK

Hasil output program dapat dilihat pada gambar di bawah:



Gambar 6. Hasil output

4. KESIMPULAN DAN SARAN

Metode *support vector machine* dengan pendekatan *one vs one* dan *one vs rest* menghasilkan hasil yang sama hal ini bergantung terhadap dataset yang digunakan. Pembagian data latih dan data uji yang proposional akan menghasilkan hasil yang baik. Model terbaik dihasilkan oleh model *one vs one* dan *one vs rest* dengan pembagian 75% data latih dan 25% data uji dengan nilai akurasi 89.58%. Streamlit merupakan *library web framework* yang menghemat waktu untuk membangun *website* dengan python. Aplikasi berhasil mempermudah proses labelling data dan membedakan jenis komentar dalam video Youtube. Saran untuk penelitian selanjutnya adalah mencoba algoritma *deep learning* untuk nilai akurasi yang lebih baik, menambah fitur untuk menghapus untuk komentar yang terdeteksi spam dan menggunakan teknik *stemming* Bahasa Indonesia yang lebih cepat dibandingkan sastrawi.

Ucapan Terima Kasih (*Acknowledgement*)
Penulis berterima kasih kepada seluruh pihak yang telah membantu proses penyelesaian penelitian ini.

REFERENSI

Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal Tekno Kompak*, *14*(2), 86-91.

Baeldung, (2022, September 25). "*Multiclass Classification using Support Vector Machine*". Baeldung https://www.baeldung.com/cs/svm-multiclass-classification

Ichwan, M., & Dewi, I. A. (2018). Klasifikasi Support Vector Machine (SVM) Untuk Menentukan Tingkat Kemanisan Mangga Berdasarkan Fitur Warna. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, *3*(2), 16-23.

https://doi.org/10.24912/jsstk.v1i2.28510

Iqbal, A. R., & Miftahuddin, Y. (2022). Implementasi SVM Untuk Deteksi Komentar Negatif Berbahasa Indonesia Di Twitter. *FTI*.

Jihad, J., Widiastuti, N. I., & Dewi, K. E. (2021). Support Vector Machine Untuk Ekstraksi Dokumen Karya Ilmiah. *Komputa: Jurnal Ilmiah Komputer Dan Informatika*, *10*(2), 87-94.

Julio, R., Pratiwi, H., & Wahyuningsih, Y. (2022). Pendeteksi Komentar Spam Youtube Menggunakan Bag Of Word Dan Random Forest. *Scientico: Computer Science And Informatics Journal*, *5(1)*, 7-14.

Prayoga, F. A., Pinandito, A., & Perdana, R. S. (2018). Rancang Bangun Aplikasi Deteksi Spam Twitter Menggunakan Metode Naive Bayes Dan Knn Pada Perangkat Bergerak Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, *2548*, 964.

Rosari, M. A., Wasino, W., & Tony, T. (2022). Analisis Sentimen Tanggapan Masyarakat Terhadap Bantuan Sosial Pemerintah Di Masa Pandemi Covid-19 Pada Platform Twitter. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *10(1).*

Syam, A. T., Fahri, A., Saputra, B., Maulana, R. F., Dwiani, S., Holid, W. G., & Firmansyah, R. (2020). Klasifikasi Komentar Spam Pada Instagram Menggunakan Metode Support Vector Machine. *Buffer Informatika*, *6(2)*, 1-5.

Thomas, V. W. D., & Rumaisa, F. (2022). Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine Dan TF-IDF. *Jurnal Media Informatika Budidarma, 6(3)*, 1767-1774.

Halaman ini sengaja dikosongkan

https://doi.org/10.24912/jsstk.v1i2.28510