

## ABSTRAK

**Dwi Raka Mudiarta, I Made** (2022), *Analisis Sentimen pada Dokumen Twitter Mengenai Vaksin COVID-19 dengan Damerau Levenshtein Distance dan Support Vector Machine*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. dan Pembimbing II: Dr. Komang Setemen, S.Si., M.T.

*Kata-kata kunci:* sentimen, analisis, twitter, support vector machine, damerau levenshtein distance

Salah satu media sosial yang paling banyak digunakan di Indonesia adalah *Twitter*, dimana Indonesia termasuk 5 besar pengguna terbesar di dunia. Pada *twitter* user dapat dengan bebas memberikan pendapat mengenai suatu topik yang sedang tren di masyarakat. Opini masyarakat terkait vaksin Covid-19 banyak dikemukakan pada layanan media sosial yang digunakan masyarakat, salah satunya pada *Twitter*. Oleh sebab itu, penting bagi pemerintah untuk mengetahui bagaimana sentimen masyarakat terkait kebijakan vaksin Covid-19 ini. Sentimen masyarakat terhadap kebijakan vaksin berdasarkan opini yang disampaikan dapat diketahui dengan menggunakan metode *text mining*. Penelitian ini dipecah menjadi beberapa tahapan yaitu 1) *preprocessing* data, 2) ekstraksi fitur, 3) klasifikasi data dan 4) pengujian hasil. Pada tahap *preprocessing* perbaikan ejaan, penulis menggunakan algoritma *Damerau Levenshtein Distance*. Ekstraksi fitur dilakukan dengan menggunakan TF dan TF-IDF. Hasil pengujian menggunakan *confusion matrix* pada SVM dengan menggunakan algoritma *Damerau Levenshtein Distance* pada tahap *text preprocessing* mendapatkan nilai *accuracy* 85,07%, *recall* positif 87,84%, *recall* negatif 82,27%, *precision* positif 83,33%, dan *precision* negatif 87,02%. Untuk hasil pengujian SVM tanpa menggunakan algoritma *Damerau Levenshtein Distance* pada tahap *text preprocessing* diperoleh nilai *accuracy* sebesar 82,58%, *recall* positif 83,33%, *recall* negatif 81,82%, *precision* positif 82,22%, dan nilai *precision* negatif sebesar 82,95%. Peningkatan hasil klasifikasi diperoleh setelah mengimplementasikan algoritma *Damerau Levenshtein Distance* untuk nilai komponen pengujian *accuracy* mengalami peningkatan sebesar 2,49%, nilai *recall* positif 4,50%, nilai *recall* negatif 0,45%, nilai *precision* positif 1,11% dan *precision* negatif 4,07%. Perbaikan kata dapat meningkatkan akurasi namun menyebabkan waktu proses yang lama karena harus membandingkan kata yang akan diperbaiki dengan kata dalam kamus satu persatu. Pada penelitian ini baru menggunakan kelas positif dan negatif, selanjutnya dapat dikembangkan dengan penambahan kelas netral serta diuji akurasinya dengan jumlah kelas yang baru dan dapat dikembangkan dengan menambahkan sumber data dari media sosial lain.

## ABSTRACT

**Dwi Raka Mudiarta, I Made** (2022), *Sentiment Analysis on Twitter Documents Regarding the COVID-19 Vaccine with Damerau Levenshtein Distance and Support Vector Machine*. Thesis, Computer Science, Postgraduate program, Ganesha University of Education.

This thesis has been supervised and approved by Supervisor I: Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. and Supervisor II: Dr. Komang Setemen, S.Si., M.T.

*Keywords:* sentiment, analysis, twitter, support vector machine, damerau levenshtein distance

Twitter is one of the most widely used social media in Indonesia, where Indonesia is among the top 5 largest users in the world. Twitter users can freely provide various comments containing opinions related to an object or topic. Public opinion regarding the Covid-19 vaccine has been expressed a lot on social media services used by the community, one of which is on Twitter. Therefore, the government needs to know how the public's sentiment regarding the Covid-19 vaccine policy is. Public sentiment on vaccine policy based on the opinions expressed can be known by using the text mining method. This research is divided into several stages, namely 1) data preprocessing, 2) feature extraction, 3) data classification, and 4) results testing. In the preprocessing stage of spelling correction, the author proposes the Damerau Levenshtein Distance algorithm. Feature extraction is done using TF and TF-IDF. The test results using the confusion matrix on SVM using the Damerau Levenshtein Distance algorithm at the text preprocessing stage get an accuracy value of 85.07%, positive recall 87.84%, negative recall 82.27%, positive precision 83.33%, and negative precision 87,02%. For the results of the SVM test without using the Damerau Levenshtein Distance algorithm at the text preprocessing stage, the accuracy value is 82.58%, positive recall is 83.33%, negative recall is 81.82%, positive precision is 82.22%, and negative precision value is 82.95%. Improved classification results were obtained after implementing the Damerau Levenshtein Distance algorithm for the accuracy test component value increased by 2.49%, positive recall value 4.50%, negative recall value 0.45%, positive precision value 1.11%, and negative precision 4,07%. Correcting words can increase accuracy but causes a long processing time because they have to compare the words to be corrected with the words in the dictionary one by one. In this study, only positive and negative classes were used, then it could be developed by adding a neutral class and testing its accuracy with a new number of classes and could be developed by adding data sources from other social media.