

## ABSTRAK

Program vaksinasi Covid-19 menimbulkan pro dan kontra dari masyarakat mengenai keamanan dari vaksin itu sendiri. Dengan banyaknya komentar masyarakat Indonesia mengenai vaksin Covid-19 tersebut maka dapat diperoleh sumber data yang dapat dimanfaatkan untuk melakukan penelitian yang bermanfaat, salah satunya untuk melakukan klasifikasi persepsi masyarakat (analisis sentimen) terhadap vaksin Covid-19. Berbagai macam algoritma telah digunakan untuk menyelesaikan permasalahan analisis sentimen yang tentunya memiliki kelebihan dan kekurangannya masing-masing. Namun hal yang masih sering terjadi yaitu dalam menangani permasalahan analisis sentimen sangat erat sekali terjadi kesalahan penulisan pada data komentar yang ada, sehingga sistem salah mengidentifikasi komentar yang mengakibatkan ketidakakuratan dalam melakukan klasifikasi.

Hal ini perlu diatasi dengan cara menambahkan metode yang digunakan saat proses *preprocessing* yaitu dengan melakukan perbaikan kata. Pada penelitian ini menggunakan algoritma *Levenshtein Distance* untuk melakukan perbaikan kata sehingga dapat mengatasi kesalahan-kesalahan penulisan kata yang terdapat pada dataset komentar masyarakat tentang vaksin Covid-19 yang berasal dari akun *Instagram* Kemenkes RI yang berjumlah 2394 data. Data yang sudah dilakukan *preprocessing* dan perbaikan kata kemudian dilakukan klasifikasi menggunakan algoritma *Naïve Bayes* untuk mengetahui hasil sentimennya.

Dalam penelitian ini, pengujian dilakukan menggunakan *confusion matrix* dengan dua skenario yaitu data yang dilakukan *preprocessing* perbaikan kata dan data yang tidak dilakukan *preprocessing* perbaikan kata. Data uji yang digunakan terbagi menjadi dua yaitu data uji lama yang berjumlah 479 data dan data uji baru yang belum pernah dilakukan training sebelumnya berjumlah 100 data. Akurasi pengujian menggunakan data uji lama yang berjumlah 479 data meningkat dari 61% menjadi 71% dan pengujian dengan data uji baru yang berjumlah 100 data akurasi meningkat dari 59% menjadi 66%. Namun untuk klasifikasi data testing baru memperoleh akurasi yang cukup rendah walaupun data yang dites hanya berjumlah 100 data, hal ini disebabkan oleh sistem yang kurang mampu dalam melakukan klasifikasi data baru yang belum pernah dilakukan training sebelumnya.

**Kata kunci** : analisis sentimen, vaksin, Naïve Bayes, perbaikan kata, Levenshtein Distance