ABSTRAK

Pandemi *Covid-19* belum berakhir di Indonesia, penyebaran *Covid-19* saat ini di Indonesia masih terus bertambah. Salah satu upaya penanganan pemerintah dalam mengurangi penyebaran *Covid-19* yaitu dengan melakukan vaksinasi. Diadakannya *vaksin* tersebut dengan tujuan yakni untuk mengatasi penyebaran virus *Covid-19*, mengurangi angka kematian dan untuk mencapai *herd immunity.* Namun masih banyak masyarakat yang belum melakukan vaksinasi, karena masyarakat masih ragu untuk melakukan vaksinasi. Banyak masyarakat yang termakan akan berita *Hoax*, sehingga menimbulkan sentimen publik salah satunya di media sosial *Twitter.* Penelitian ini bertujuan untuk melakukan perbandingan hasil klasifikasi antara penggunaan metode *Modified K-Nearest Neighbor* dan metode *Levenshtein Distance* dengan penambahan seleksi fitur *Chi Square* dalam melakukan analisis sentimen publik dari *tweet* opini masyarakat mengenai *vaksin*. Data yang digunakan diambil dari *Twitter API* dengan kata kunci "*vaksin*" dengan jumlah data sebanyak 5000. Data memiliki label positif dan negatif setelah melalui tahapan *labeling* yang menggunakan *tools VADER Lexicon.* Berdasarkan hasil yang diperoleh menunjukan bahwa algoritma *Levenshtein Distance* dapat dengan baik membantu dalam melakukan klasifikasi yaitu dengan memperbaiki kata-kata yang tidak sesuai atau kata yang *typo*. Peningkatan akurasi lebih dominan terhadap penggunaan kombinasi algoritma *MKNN, Levenshtein Distance* dan seleksi fitur *Chi Square,* bila dibandingkan dengan penggunaan kombinasi algoritma *MKNN* dengan seleksi fitur *Chi Square.* Peningkatan nilai akurasi tersebut disebabkan oleh adanya kata yang typo dinormalisasikan sehingga dapat menambah frekuensi kata yang ada. Akurasi terbaik diperoleh pada k = 7 dan nilai *k-fold* = 9 dengan akurasi yang diperoleh sebesar 75.09% pada penggunaan algoritma *Levenshtein Distance,* sedangkan hasil akurasi tanpa menggunakan *Levenshtein Distance* pada nilai *k-fold* 9 dan jumlah tetangga (k) 7 sebesar 72.02%.

Kata Kunci : *Vaksin, Twitter API, Modified K-Nearest Neighbor, Chi Square, Levenshtein Distance*

ABSTRACT

Pandemic *Covid-19* has not ended in Indonesia, the current spread of *Covid-19* in Indonesia is still growing. One of the government's efforts to reduce the spread of *Covid-19* is by vaccinating. *Covid vaccine -19* virus, reduce mortality and achieve *herd immunity.* However, there are still many people who have not vaccinated, because people are still hesitant to vaccinate. Many people are consumed by *hoax*, causing public sentiment, one of which is on social media *Twitter.* This study aims to compare the classification results between the use of the *Modified K-Nearest Neighbor* method and the *Levenshtein Distance* with the addition of *Chi Square* in analyzing public sentiment from public *tweets* about *vaccines*. The data used is taken from *the Twitter API* with the keyword "*vaccine*" with a total of 5000 data. The data has positive and negative labels after going through *labeling* using *Lexicon's VADER tools.* Based on the results obtained, it shows that the *Levenshtein Distance* can properly assist in classifying, namely by correcting words that are not appropriate or words that are *typo*. The increase in accuracy is more dominant with the use of a combination of the *MKNN algorithm, Levenshtein Distance* feature selection *Chi Square* when compared to the use of a combination of the *MKNN* feature selection *Chi Square.* The increase in the accuracy value is caused by the presence of normalized typo words so that it can increase the frequency of existing words. The best accuracy is obtained at k = 7 and the value *of k-fold* = 9 with an accuracy obtained of 75.09% using the *Levenshtein Distance algorithm,* while the accuracy results without using *Levenshtein Distance* at a *k-fold* 9 and the number of neighbors (k) 7 is 72.02%.

Keywords : *Vaccine, Twitter API, Modified K-Nearest Neighbor, Chi Square, Levenshtein Distance.*