

ABSTRAK

Polymerase Chain Reaction (PCR) merupakan metode diagnosis pemeriksaan virus *Covid-19* dengan cara mendeteksi DNA virus. Pemerintah menerapkan kebijakan tes PCR sebagai syarat *Skrining* bagi masyarakat yang ingin melakukan perjalanan. Namun, kebijakan tes PCR tersebut selalu berubah-ubah dengan cepat, mulai dari pemberlakuan, masa berlaku, dan tarif. Hal tersebut tentunya menimbulkan kebingungan pada masyarakat yang dapat dituangkan pada media *Twitter*. Penelitian ini bertujuan untuk melakukan perbandingan hasil klasifikasi antara penggunaan metode *Support Vector Machine* (SVM) tanpa dan dengan *Levenshtein Distance*, mengetahui pengaruh penggunaan *Levenshtein Distance* terhadap hasil klasifikasi, dan mencari akurasi optimal dari SVM dalam melakukan analisis sentiment *Twitter* terhadap PCR. Data yang digunakan diambil dari *Twitter API* dengan kata kunci “tes pcr” dengan jumlah data yang berhasil diperoleh sebanyak 2078. Data menjadi berjumlah 1088 setelah melalui proses penghilangan data yang banyak mengandung bahasa asing dan data yang mengalami *Missing Value*. Data memiliki label positif dan label negatif setelah melalui tahapan pengelompokan menggunakan *Vader*. Berdasarkan hasil pengujian yang telah dilakukan, akurasi tertinggi terdapat pada pengujian SVM dengan *Levenshtein Distance*, dengan *kernel Polynomial*, dengan parameter *C* bernilai 1 dan parameter *Gamma* bernilai 1 dengan akurasi 90,73%. Pengaruh pada penggunaan *Levenshtein Distance* terlihat pada terdapatnya pengujian yang mengalami kenaikan dan penurunan akurasi. Selain itu, akurasi juga dipengaruhi oleh penggunaan *kernel Rbf* dan *kernel Polynomial* yang sangat bergantung pada parameter yang digunakan dalam penelitian ini. Untuk akurasi SVM yang optimal didapatkan pengujian dengan skema menggunakan *Levenshtein Distance*, dengan *kernel Polynomial*, dan parameter *C* bernilai 1 dan parameter *Gamma* bernilai 1, dengan akurasi 90,73%, sedangkan bila tidak menggunakan *Levenshtein Distance* 90,51%.

Kata Kunci: *Tes PCR, Twitter API, Support Vector Machine, Levenshtein Distance.*

ABSTRACT

Polymerase chain reaction (PCR) is a diagnostic method for examining the Covid-19 virus by detecting DNA virus. The government has implemented a PCR test policy as a screening requirement for people who want to travel. However, the PCR test policy is always changing rapidly, starting from its implementation, validity period, and tariffs. This of course causes confusion in the community which can be poured on *Twitter* social media. This study aims to compare the results of the classification between the use of the *Support Vector Machine* (SVM) method without and with *Levenshtein Distance*, the effect of using *Levenshtein Distance* on the classification results, and find the optimal accuracy of SVM in analyzing *Twitter* sentiment against PCR. The data used was taken from the *Twitter API* with the keyword “pcr test” with the amount of data that was successfully obtained as much as 2078. The data became 1088 after going through the process of eliminating data that contained many foreign languages and data that had *missing values*. The data has a positive label and a negative label after going through the sentiment grouping stages using *Vader*. Based on the results of the tests that have been carried out, the highest accuracy is found in the SVM test with *Levenshtein Distance*, with a *Polynomial Kernel*, with parameter *C* of 1 value and *Gamma* parameter of 1 with an accuracy of 90.73%. The effect on the use of *Levenshtein Distance* is seen in the presence of tests that increase and decrease in accuracy. In addition, accuracy is also affected by the use of the *Rbf kernel* and *Polynomial kernel* which are very dependent on the parameters used in this study. For optimal SVM accuracy, it is obtained when *testing* the scheme using *Levenshtein Distance*, with a *Polynomial kernel*, and the parameter *C* is worth 1 and the *Gamma* parameter is 1, with an accuracy of 90.73%, whereas if not using *Levenshtein Distance* 90,51%.

Keywords: *PCR Test, Twitter API, Support Vector Machine, Levenshtein Distance.*