

## ABSTRAK

Kekerasan dan pelecehan seksual adalah topik yang terus menarik perhatian kita dalam pembicaraan sehari-hari. Media sosial, termasuk Twitter, memberikan wadah bagi pengguna untuk saling berinteraksi dan berbagi informasi. Namun, sering kali Twitter disalahgunakan sebagai sarana untuk melakukan tindakan seperti *cyberstalking*, *cyberbullying*, dan pelecehan seksual. Pelecehan seksual di platform ini dapat berupa komentar-komentar yang tidak pantas secara seksual atau bahkan lelucon yang kasar. Penelitian ini bertujuan untuk mengetahui unjuk kerja penggunaan variasi nilai K pada *K-Nearest Neighbors*, *K-Fold* dan variasi K *Chi-Square* dalam melakukan analisis sentimen terhadap *tweet* tentang Pelecehan Seksual untuk mendapatkan akurasi optimal. Proses melabeli data set mengandalkan kamus *Sentistrength* yang dikembangkan oleh *CyberEmotion*. Sementara itu, variasi nilai K pada *K-Nearest Neighbors*, *K-Fold* dan variasi K *Chi-Square* digunakan untuk mendapatkan akurasi optimal dalam melakukan analisis sentimen terhadap *tweet* tentang Pelecehan Seksual untuk mendapatkan akurasi optimal. Data yang digunakan terdiri dari label positif, negatif, dan netral yang diperoleh melalui *Twitter API* menggunakan bantuan *Rapid Miner* dengan jumlah data 10996 data. Adapun hasil penelitian menunjukkan variasi nilai K pada *K-Nearest Neighbors*, *K-Fold* dan variasi K *Chi-Square* yang digunakan untuk mendapatkan akurasi optimal berdasarkan hasil pengujian yang dilakukan dengan akurasi tertinggi yang diperoleh yaitu 80.47% ada pada variasi K *K-Nearest Neighbors*= 7, *K-Fold*= 11, dan K *Chi-Square*= 200.

**Kata Kunci:** Pelecehan Seksual, *Twitter API*, Analisis Sentimen, *Sentistrength*, *K-Nearest Neighbor*, *Chi-Square*, *K-Fold*

## ABSTRACT

Sexual violence and harassment are topics that continue to catch our attention in our daily conversations. Social media, including Twitter, provides a platform for users to interact with each other and share information. However, Twitter is often misused as a means to commit acts such as cyberstalking, cyberbullying and sexual harassment. Sexual harassment on this platform can take the form of sexually inappropriate comments or even crude jokes. This study aims to determine the performance of using variations of K values in K-Nearest Neighbors, *K-Fold* and variations of K Chi-Square in performing sentiment analysis of tweets about Sexual Harassment to get optimal accuracy. The process of labeling the data set relies on the Sentistrength dictionary developed by CyberEmotion. Meanwhile, the variation of K value in K-Nearest Neighbors, *K-Fold* and K *Chi-Square* variation is used to get optimal accuracy in conducting sentiment analysis on tweets about Sexual Harassment to get optimal accuracy. The data used consists of positive, negative, and neutral labels obtained through the Twitter API using Rapid Miner with the help of 10996 data. The results showed variations in the K value of K-Nearest Neighbors, *K-Fold* and K Chi-Square variations used to get optimal accuracy based on the results of the tests carried out with the highest accuracy obtained, namely 80.47% in the variation of K K-Nearest Neighbors = 7, *K-Fold* = 11, and K Chi-Square = 200.

**Keywords:** Sexual Harassment, Twitter API, Sentiment Analysis, Sentistrength, K-Nearest Neighbors, Chi-Square, *K-Fold*