

## INTISARI

Jakarta menghadapi masalah serius polusi udara yang berdampak negatif pada kesehatan dan lingkungan. Penelitian ini menggunakan analisis sentimen untuk mengevaluasi opini publik terhadap polusi udara di Jakarta. Dengan tujuan untuk membandingkan metode SVM dan CNN pada analisis sentimen komentar pengguna Twitter terhadap polusi udara di Jakarta menggunakan dataset yang berjumlah 1439 data yang diperoleh dengan teknik crawling menggunakan kata kunci “polusi udara Jakarta” dan “polusi jakarta”. Data diolah pada preprocessing data dan pelabelan data menggunakan VADER Lexicon menghasilkan 1038 sentimen positif dan 401 sentimen negatif. Pengujian menggunakan metode Support Vector Machine (SVM) dan Convolution Neural Network (CNN) dilakukan dengan dua skenario pembagian data training dan testing yaitu, rasio 80:20 dan 90:10. Hasil dari penelitian ini yaitu, metode SVM memiliki performa yang lebih baik dibandingkan dengan metode CNN. Hasil akurasi terbaik metode SVM dengan kernel linear diperoleh dari pengujian dengan rasio 90:10 pada parameter Cost bernilai 1 memperoleh nilai akurasi sebesar 88,89%. Sedangkan metode CNN memperoleh nilai akurasi sebesar 88,19% pada nilai epoch sebesar 9.

Kata kunci: analisis sentimen, SVM, CNN, Twitter, polusi udara Jakarta

## **ABSTRACT**

*Jakarta faces a serious problem of air pollution which has a negative impact on health and the environment. This research uses sentiment analysis to evaluate public opinion towards air pollution in Jakarta. With the aim of comparing the SVM and CNN methods in sentiment analysis of Twitter user comments on air pollution in Jakarta using a dataset of 1439 data obtained by crawling techniques using the keywords "Jakarta air pollution" and "Jakarta pollution". The data was processed in data preprocessing and data labeling using VADER Lexicon resulting in 1038 positive sentiments and 401 negative sentiments. Testing using the Support Vector Machine (SVM) and Convolution Neural Network (CNN) methods was carried out with two scenarios for dividing training and testing data, namely, a ratio of 80:20 and 90:10. The results of this research are that the SVM method has better performance compared to the CNN method. The best accuracy results for the SVM method with a linear kernel were obtained from testing with a ratio of 90:10 with a Cost parameter of 1, obtaining an accuracy value of 88.89%. Meanwhile, the CNN method obtained an accuracy value of 88.19% at an epoch value of 9.*

*Keywords: sentiment analysis, SVM, CNN, Twitter, Jakarta air pollution*