

ABSTRAK

Pada bulan Mei 2023, Indonesia dipercayakan untuk menjadi tuan rumah event Internasional yaitu *Konferensi Tingkat Tinggi ASEAN* yang diadakan di Labuan Bajo, Nusa Tenggara Timur. Konferensi Tingkat Tinggi ASEAN sendiri merupakan pertemuan tingkat tinggi para kepala negara atau pemerintahan negara anggota yang membahas isu – isu strategis dan memperkuat kerja sama regional. Hal ini memicu banyak pendapat publik dari kalangan masyarakat yang diutarakan melalui media sosial *X*. Oleh karena itu akan sangat membantu apabila dilakukan suatu *sentiment analysis* sebagai salah satu cara untuk membantu *pemerintah* dalam mengevaluasi masukan dari masyarakat dengan cara mengklasifikasi *tweet* menjadi positif, negatif, dan netral. Data yang digunakan diambil menggunakan metode *web scraping* dengan menggunakan *python* sebanyak 13.431 *tweet*. *Tweet* tersebut berisi informasi – informasi yang dapat digunakan untuk mengetahui sentimen masyarakat terhadap topik penelitian. Data tersebut akan diberi label yang terdiri dari *positif*, *negatif*, dan *netral* menggunakan *lexicon based*. Penelitian ini menggunakan metode klasifikasi *K-Nearest Neighbors* untuk mendapatkan akurasi yang terbaik. Hasil akurasi paling optimal yang dapat dihasilkan oleh algoritma *K-Nearest Neighbors* dengan menggunakan *seleksi fitur chi – square* adalah 80,2219% dengan *precision* 80,2215%, dan *recall* 80,9712% pada kondisi 11 *k – fold* dan jumlah *k – tetangga* adalah 3. Sedangkan hasil akurasi paling optimal yang dapat dihasilkan tanpa menggunakan *seleksi fitur chi – square* adalah 78,7572% dengan *precision* 78,7570%, dan *recall* 80,7001% pada kondisi 11 *k – fold* dan jumlah *k – tetangga* adalah 3. Berdasarkan hasil perbandingan diatas maka dapat disimpulkan bahwa pada percobaan klasifikasi menggunakan metode KNN dengan *seleksi fitur skor chi – square* lebih baik dibandingkan dengan menggunakan metode KNN tanpa *seleksi fitur skor chi – square* dalam mengklasifikasi komentar Masyarakat di *X* mengenai Konferensi Tingkat Tinggi ASEAN di Indonesia.

Kata Kunci : *Konferensi, ASEAN, Regional, Sentiment Analysis, Tweet, Web Scraping, Python, VADER, K-Nearest Neighbors, TF-IDF, K-Fold Cross Validation, Akurasi, Precision, Recall, Chi – Square.*

ABSTRACT

In May 2023, Indonesia was entrusted to host an international event, the ASEAN Summit held in Labuan Bajo, East Nusa Tenggara. The ASEAN Summit itself is a high-level meeting of heads of state or government of member countries that discuss strategic issues and strengthen regional cooperation. This triggered many public opinions from the community expressed through the social media X. Therefore, it would be very helpful if a sentiment analysis is carried out as a way to assist the government in evaluating input from the public by classifying tweets into positive, negative, and neutral. The data used was taken using the web scraping method using python as many as 13,431 tweets. The tweets contain information that can be used to determine public sentiment towards the research topic. The data will be labeled consisting of positive, negative, and neutral using lexicon based. This research uses the K-Nearest Neighbors classification method to get the best accuracy. The most optimal accuracy result that can be generated by the K-Nearest Neighbors algorithm using chi-square feature selection is 80.2219% with a precision of 80.2215%, and a recall of 80.9712% in conditions of 11 k-fold and the number of k-neighbors is 3. While the most optimal accuracy result that can be generated without using chi-square feature selection is 78.7572% with a precision of 78.7570%, and a recall of 80.7001% in conditions of 11 k-fold and the number of k-neighbors is 3. Based on the comparison results above, it can be concluded that the classification experiment using the KNN method with chi-square score feature selection is better than using the KNN method without chi-square score feature selection in classifying public comments on X regarding the ASEAN Summit in Indonesia.

Keywords : *Konferensi, ASEAN, Regional, Sentiment Analysis, Tweet, Web Scraping, Python, VADER, K-Nearest Neighbors, TF-IDF, K-Fold Cross Validation, Akurasi, Precision, Recall, Chi – Square.*