

ABSTRAK

Dunia teknologi sekarang ini sedang mengalami perkembangan yang cukup signifikan, masyarakat kini bisa dengan bebas mengakses dan memanfaatkan media sosial sebagai contoh yakni Twitter. Banyak sekali pengguna Twitter mencurahkan pendapat melalui tweet-tweet yang mereka kirimkan pada media sosial tersebut, khususnya mengenai Pekan Olahraga Nasional (PON) XX 2021 silam. Banyak sekali tweet yang bersifat mendukung, tetapi tidak jarang juga ada tweet yang bersifat keluhan mengenai penyelenggaraan PON XX tersebut. Dari masalah tersebut, dilaksanakanlah penelitian mengenai analisis sentimen pada data twitter yang berkaitan dengan PON XX dan mempergunakan metode Naïve Bayes. Jumlah data yang di crawling sebanyak 1000 data lalu data melalui proses *dataduplicate removal* sehingga memperoleh hasil data sebanyak 218 data tweet dan belum terlabelkan. Sebelum proses mengklasifikasikan data yang diperoleh, data harus dilaksanakan pemberian label pada datanya serta pembersihan data terlebih dulu sebelum masuk pada tahapan *text processing*, selanjutnya data diberi bobot pada tiap kata dengan *Term Frequency– Inverse Document Frequency* (TF-IDF) yang akan kedepannya kata tersebut akan dijadikan sebagai fitur. Lalu, karena data berlabel positif serta negatif mempunyai jumlah yang jauh berbeda, maka dimanfaatkan metode *Synthetic Minority Oversampling Technique* (SMOTE) guna melaksanakan penyeimbangan terhadap datanya. Tahapan selanjutnya dilaksanakan pembagian data yang besarnya yakni 80% 20% dan diklasifikasikan dengan metode Naive Bayes. Hasil yang diperoleh dari pelaksanaan penelitiannya tersebut ialah diperoleh bahwa data uji memperoleh accuracy yang besaran persentasenya yakni 99%, precision dengan besaran persentasenya 100%, recall dengan besaran persentasenya 98%.

Kata Kunci: Analisis Sentimen, Klasifikasi, PON XX, Naïve Bayes

CLASSIFICATION AND SENTIMENT ANALYSIS ON TWITTER DATA USING NAÏVE BAYES. (CASE STUDY: PEKAN OLAHRAGA NASIONAL XX 2021)

Krisna Jonathan Sitorus

ABSTRACT

The world of technology is currently developing very rapidly, people can now freely access and express themselves using social media as an example, namely Twitter. Many Twitter users express their opinions through the tweets they send on social media, especially regarding the XX 2021 National Sports Week (PON) ago. Several types of tweets that are appreciative of the event are often seen, but not infrequently there are also tweets that are complaining about the implementation of the XX PON. From this problem, a research was conducted on sentiment analysis on twitter data regarding PON XX using the Naïve Bayes method. The amount of data that is crawled is 1000 data and then the data goes through a duplicate data removal process so that it produces 218 tweets and has not been labeled. Before the process of classifying the data obtained, the data must be labeled and cleaned first before entering the text processing stage, then the data is given a weight for each word with a Term Frequency–Inverse Document Frequency (TF-IDF) which will be used in the future as a feature. Then, because the data with positive and negative labels have significantly different amounts, the Synthetic Minority Oversampling Technique (SMOTE) method was used to balance the data. The next stage is the distribution of data by 80% 20% and classified by the Naive Bayes method. The results obtained from this study are obtained that the test data get 99% accuracy, 100% precision, 98% recall.

Keyword: Sentiment Analysis, Classification, PON XX, Naïve Bayes