

ABSTRAK

Ayu Juli Astari, Ni Made (2020), *Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I : Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. dan Pembimbing II: Dr. Gede Indrawan, S.T., M.T.

Kata-kata kunci: analisis sentimen, dampak corona, klasifikasi, naive bayes, twitter.

Awal tahun 2020 Indonesia mengalami wabah oleh virus Corona dan banyak warga Indonesia yang terkena dampak penularan virus ini. Pemerintah mengeluarkan kebijakan pembatasan aktivitas keluar rumah dengan tujuan mengurangi jumlah penderita di Indonesia. Tidak sedikit opini yang muncul pada sosial media dengan diberlakukannya kebijakan tersebut. Diperlukan analisis sentimen untuk mengetahui kecenderungan sentimen opini yang muncul pada sosial media. Tujuan dari penelitian ini adalah memperoleh analisis dokumen teks untuk mendapatkan sentimen positif atau negatif dari opini. Data yang digunakan merupakan dokumen *tweet* dari *Twitter* mengenai dampak virus Corona yang kemudian diberi label sentimen positif atau negatif. Jumlah data yang digunakan sebanyak 796 dokumen *tweet*. Dokumen dengan sentimen positif sebanyak 290 dokumen dan dokumen dengan sentimen negatif sebanyak 506 dokumen. Data yang terkumpul dibagi untuk digunakan sebagai data latih dan data testing untuk proses klasifikasi. Tahapan yang dilakukan yaitu pengumpulan data, *preprocessing*, klasifikasi, dan evaluasi. Metode yang digunakan dalam penelitian ini adalah Metode Naive Bayes Classifier untuk mengklasifikasi dokumen *tweet* dengan pertimbangan tingginya *accuracy* hasil klasifikasi pada penelitian sebelumnya. Hasil klasifikasi berupa sentimen positif dan negatif dari opini. Hasil klasifikasi dievaluasi menggunakan *accuracy* dan *error rate* dengan tujuan mengetahui keakuratan dokumen setelah diklasifikasi menjadi sentimen positif atau negatif. Hasil pengujian tertinggi yaitu *accuracy* sebesar 67% dan *error rate* sebesar 33%. Saran untuk penelitian selanjutnya adalah dengan memperhitungkan emoji yang digunakan dalam *tweet*, penambahan kosa kata yang terkini karena meningkatnya jumlah kosa kata baru dalam sosial media, dan mempertimbangkan unsur semantik kata yang digunakan pada opini.

ABSTRACT

Ayu Juli Astari, Ni Made (2020), *Sentiment Analysis on Twitter Documents in Corona Virus Impact using Naive Bayes Classifier*. Thesis, Computer Science, Post Graduate Study Program, Ganesha University of Education.

This thesis has been supervised and approved by Supervisor I : Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. and Supervisor II: Dr. Gede Indrawan, S.T., M.T.

Keywords: sentiment analysis, corona impact, classification, naive bayes, twitter.

Indonesia contracted the Corona virus in early 2020 and many residents were affected by this virus. The government issued a policy of restricting outdoor activities with the aim of reducing the number of sufferers in Indonesia. Many opinions appear on social media with the enactment of this policy. Sentiment analysis is needed to determine the trend of opinion sentiment that appears on social media. The purpose of this study was to obtain an analysis of text documents in the form of positive or negative sentiments of opinion. The data used is a tweet document from Twitter regarding the impact of the Corona virus which is then labeled as positive or negative sentiment. The amount of data used was 796 tweet documents. There were 290 positive documents and 506 negative documents. The collected data is divided to be used as training data and testing data for the classification process. The stages in this research are document collection, preprocessing, classification, and evaluation. The method used in this study is the Naive Bayes Classifier to classify tweet documents by considering the high accuracy of the classification results in previous studies. The results of the classification are in the form of positive and negative sentiments of opinion. The classification results are evaluated using the accuracy and error rate with the aim of knowing the accuracy of the document after being classified into positive or negative sentiments. The highest test result is the accuracy of 67% and the error rate of 33%. Suggestions for further research are to take into account the emojis used in tweets, the addition of the latest vocabulary due to the increasing number of new vocabulary in social media, and to consider the semantic elements of words used in opinions.

UNDIKSHA