

## **KLASIFIKASI DOKUMEN KOMENTAR PADA SITUS YOUTUBE MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

**MOH AZIZ NUGROHO**

(Pembimbing : Heru Agus Santoso, Ph.D)

*Teknik Informatika - S1, FIK, Universitas Dian Nuswantoro*

*www.dinus.ac.id*

*Email : 111201105980@mhs.dinus.ac.id*

### **ABSTRAK**

Di era serba maju seperti ini persaingan dalam produksi film sangatlah ketat, kemajuan teknologi menjadikan konsumen dapat dengan mudah mendapatkan informasi tentang film yang mereka inginkan misalkan pada situs youtube. Rumah produksi film dituntut untuk dapat memproduksi film dalam waktu singkat dan jumlah banyak agar tidak tertinggal oleh rumah produksi film lainnya karena konsumen menginginkan sebuah film yang selalu baru setiap saat. Kemudian masalah timbul karena rumah produksi film tersebut lebih mementingkan jumlah film yang diproduksi dibandingkan dengan kualitas film tersebut. Sistem klasifikasi komentar diharapkan dapat membantu untuk mengetahui respon positif dan negatif dari pengguna situs youtube yang memberikan komentarnya. Sentiment analysis digunakan untuk mengetahui sikap seseorang dalam konteks dokumen. Sentiment analysis memiliki tahapan preprocessing yang terdiri dari case folding, stopword removal, cleansing, tokenizing, dan stemming. Pembobotan kata yang digunakan adalah term frequency-inverse document frequency dan perhitungan similaritasnya menggunakan cosine similarity kemudian menggunakan K-Nearest Neighbor (K-NN) sebagai metode klasifikasinya. Hasil yang didapatkan dari implementasi metode k-nn ini cukup baik dengan uji coba sebanyak 6 kali. Rata-rata accuracy tertinggi adalah 0.806 dengan recall 0.848, precision 0.788, dan error rate 0.150. Sedangkan accuracy terendah adalah 0.669 dengan recall 0.88, precision 0.177, dan error rate 0.331.

Kata Kunci : Sentiment Analysis, K-Nearest Neighbor

## **DOCUMENT COMMENTARY CLASSIFICATION IN YOUTUBE SITE USING K-NEAREST NEIGHBOR (K-NN)**

**MOH AZIZ NUGROHO**

(Lecturer : Heru Agus Santoso, Ph.D)

*Bachelor of Informatics Engineering - S1, Faculty of Computer  
Science, DINUS University*

*www.dinus.ac.id*

*Email : 111201105980@mhs.dinus.ac.id*

### **ABSTRACT**

In this modern era, competition in movie production is very tight. Technological advance has helped consumers to easily acquire their desired movie information, such as in a site like Youtube. Production houses are required to be able to produce a large quantity of movies in quite a short time in order to keep them from being left behind by other production houses. This is due to the need of consumers who want a new movie every time. Problem then arises because the production houses tend to give more concern on movies quantity than movies quality. Comments classification system is expected to help determining both positive and negative responds from Youtube users who provide their comments. Sentiment analysis is utilized to recognize someone's attitude in the document context. Sentiment analysis has a preprocessing stage consisting of case folding, stopword removal, cleansing, tokenizing, and stemming. Words weighting used is the term frequency - inverse document frequency and similarity calculation using the cosine similarity then using K-Nearest Neighbor (K-NN) as a method of classification. Result obtained from the implementation of K-NN method is good enough with as many as six times testing. The average of highest accuracy is 0.806 with as much as 0.848 of recall, 0.788 of precision, and 0.150 of error rate. While the lowest accuracy is 0.669 with as much as 0.88 of recall, 0.177 of precision and error rate of 0.331.

Keyword : Sentiment Analysis, K-Nearest Neighbor