

## ABSTRAK

Jackie (03082180006)

### IMPLEMENTASI ALGORITMA TF-IDF DALAM KOLOM KOMENTAR AKUN *YOUTUBE CRETIVOX* UNTUK MENDETEKSI UJARAN KEBENCIAN

(xiii + 37 halaman: 24 gambar, 2 tabel)

YouTube adalah sebuah *website* penyedia layanan berbagi video secara daring yang dapat diakses oleh seluruh orang dan dapat memberikan komentar pada video yang tersedia pada kolom komentar. Akibat dari kebebasan berkomentar yang dapat dilakukan pada YouTube, orang-orang menjadi kurang mengerti batasan apa saja yang perlu dilakukan saat memberikan komentar. Tidak jarang ditemukan ujaran kebencian pada kolom komentar sebuah video. Hanya saja, masih sedikit dari kita yang dapat membedakan mana yang dapat dikategorikan dengan ujaran kebencian dan mana yang bukan. Dengan menggunakan data mining, masyarakat berpotensi untuk mengembangkan suatu sistem pendeteksian ujaran kebencian untuk mengurangi jumlah ujaran kebencian yang ditampilkan pada kolom komentar YouTube serta dapat membedakan komentar yang terindikasi ujaran kebencian dan yang bukan. Metode yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Support Vector Machine* (SVM). Langkah awal yang dilakukan adalah mengumpulkan komentar-komentar yang kemudian akan dilabeli secara mandiri. Selanjutnya data akan melewati proses *text preprocessing* dan pembobotan menggunakan metode TF-IDF sebelum diimplementasikan algoritma SVM untuk mendeteksi ujaran kebencian. Dari analisis yang telah dilakukan, metode pembobotan TF-IDF dengan implementasi algoritma SVM dengan menggunakan 80% data *training* dan 20% data *testing* berhasil menghasilkan tingkat akurasi sebesar 99% dengan nilai *precision* sebesar 100%, *f1-score* sebesar 73,68% dan *recall* sebesar 58,3%.

**Kata kunci :** *Web Scrapping*, Ujaran Kebencian, TF-IDF, SVM

Referensi : 10 (2010 - 2022)

## ABSTRACT

Jackie (03082180006)

### ***IMPLEMENTATION OF TF-IDF ALGORITHM ON COMMENT SECTION OF CRETIVOX YOUTUBE ACCOUNT TO DETECT HATE SPEECH***

*(xiii + 37 pages : 24 figures; 2 table)*

*YouTube is an online video sharing services website that can be accessed by everyone and leave remarks in the comment section. Because of the freedom of speech available on YouTube, many people are unaware of the limitations that must be adhered to when leaving comments. It's not uncommon to come across hate speech in a video's comments area. However, there are still a few of us who can tell the difference between hate speech and non-hate speech. The public has the ability to construct a hate speech detection system utilizing data mining to limit the amount of hate speech displayed in the YouTube comments section and discriminate between hate speech and non-hate speech comments. The methods used are Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machine (SVM). The initial stage is to gather comments, which will be labeled separately afterwards. Before the SVM algorithm is used to detect hate speech, the data will be preprocessed and weighted using the TF-IDF approach. From the analysis that has been carried out, the TF-IDF weighting method with the implementation of the SVM algorithm using 80% training data and 20% testing data succeeded in producing an accuracy rate of 99% with a precision value of 100%, f1-score of 73.68% and recall by 58.3%.*

**Keywords :** *Web Scrapping, Hate Speech, TF-IDF, SVM*

Reference 10 : (2010 – 2022)