

ABSTRAK

Nama : Albert
NIM : 41518120026
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Analisis Topik Dan Perbandingan Klasifikasi Pada Kolom Komentar Video Youtube Edukasi Indonesia Menggunakan Pendekatan Latent Dirichlet Allocation
Pembimbing : Achmad Kodar, Drs. MT

Ada berbagai kanal YouTube bertema edukasi yang berasal dari Indonesia dengan topic yang beragam. Namun untuk menentukan kualitas kanal tersebut, dilakukannya penelitian berdasarkan hubungan kolom komentar, judul dan video. Tujuan penelitian ini adalah mencoba untuk menemukan fokus tema dari kanal bertema edukasi yang telah dipilih untuk menjadi objek penelitian apakah kanal ini memiliki konten yang bermanfaat atau berkaitan dengan judul yang ada. Penelitian ini menggunakan data berdasarkan hasil scraping komentar dari situs YouTube menggunakan Python. Untuk melancarkan proses, dilakukan pre-processing data untuk merapikan data hasil komentar dari audiens. Algoritma SVM dan Decision Tree digunakan untuk eksperimen terhadap prediksi mengenai analisa sentimen. Latent Dirichlet Allocation merupakan metode yang handal untuk mengelompokan maupun meringkas sebuah teks yang berukuran besar. Lalu komentar tersebut dijadikan token dan diubah ke bentuk vektor dengan metode Bag of Words untuk dijadikan corpus. Setelah membuat model LDA, maka topik akan dimunculkan. Lalu ditentukan perhitungan nilai coherence untuk menentukan besaran kecocokan potongan kata dengan topik yang dihasilkan. Setelah itu ditentukan jumlah topik yang optimal dari setiap kanal untuk mencari nilai coherence yang lebih tinggi. LDA menghasilkan pembagian topik yang merata.

Kata Kunci: Natural Language Processing, Support Vector Machine, Decision Tree, Latent Dirichlet Allocation, LDA.

ABSTRACT

Name : Albert
NIM : 41518120026
Study Program : Information Technology
Title Thesis : Topic Analysis and Classification Comparison on
Youtube Comment Section of Indonesian
Education Content Creator Using Latent Dirichlet
Allocation Approaches
Counsellor : Achmad Kodar, Drs. MT

There are numerous Indonesian YouTube channels with educational themes and a range of topics. However, a study based on the correlation between the comments column, the title, and the video was done to assess the channel's quality. The goal of this research is to determine whether the educational-themed channel that has been chosen as the study's subject has a relevant emphasis and whether it is connected to any current titles. Data from YouTube comments that were scraped using Python are used in this investigation. Data pre-processing was done to clean up the data obtained from audience remarks in order to speed up the procedure. We test predictions for sentiment analysis using the SVM and Decision Tree algorithms. A dependable technique for grouping or summarizing a huge text is Latent Dirichlet Allocation. The Bag of Words approach is then used to turn the comments into a corpus by using them as tokens and vectorizing them. The topic will show up following the creation of the LDA model. The amount of match between the word fragment and the generated subject was then calculated using the coherence value. The best combination of subjects from each channel is then determined in order to get a better coherence value. An even distribution of subjects is produced using LDA.

Keywords : Natural Language Processing, Support Vector Machine, Decision Tree, Latent Dirichlet Allocation, LDA.