

# Analisis Fitur-Fitur Yang Mempengaruhi Jumlah *Subscribers* Youtube Menggunakan Algoritma *Naive Bayes Classifier*

Meliyana R. Yoanita<sup>1</sup>, Hendry Setiawan<sup>2</sup>, Paulus Lucky Tirma Irawan<sup>3</sup>

<sup>1,2,3</sup>Universitas Ma Chung

<sup>1</sup>311510019@student.machung.ac.id, <sup>2</sup>hendry.setiawan@machung.ac.id, <sup>3</sup>paulus.lucky@machung.ac.id

## ABSTRAK

Salah satunya teknologi dalam bidang komunikasi adalah platform media sosial. Media sosial Youtube adalah salah satu media sosial yang paling banyak digunakan di Indonesia. Keuntungan yang didapatkan oleh Content Creator atau Youtuber berasal dari AdSense. Youtube memiliki beberapa fitur yang disediakan yaitu like, dislike, view dan komentar (komentar dengan sentimen negatif atau positif). Diperlukan sistem klasifikasi otomatis sentimen komentar Youtube untuk klasifikasi komentar positif dan komentar negatif, selain itu analisis fitur-fitur yang mempengaruhi jumlah subscribers sehingga Content Creator dapat mengetahui fitur-fitur yang dapat mempengaruhi jumlah subscribers. Dalam penelitian ini akan dibuat sistem klasifikasi otomatis sentimen komentar menggunakan algoritma Naive Bayes (NB) sehingga proses klasifikasi komentar positif dan negatif dapat dilakukan dengan mudah, data yang digunakan dalam analisis adalah sebanyak 53 channel Youtube dengan jenis vlog. Selain itu data yang digunakan sebagai data latih klasifikasi adalah sebanyak 4166 sentimen positif dan 4166 sentimen negatif, setelah itu dilakukan analisis fitur-fitur yang mempengaruhi jumlah subscribers dengan menggunakan chi square. Hasil dari analisis dengan chi square didapatkan ada 4 fitur yang memiliki pengaruh terhadap jumlah subscribers yaitu jumlah view dengan nilai chi square 23.105, dislike dengan nilai chi square 13.745, jumlah komentar sentimen positif dengan nilai chi square 18.123 dan jumlah like dengan nilai chi square 13.745. Hasil akurasi untuk sistem klasifikasi otomatis menggunakan Naive Bayes (NB) sebesar 81%.

**Kata Kunci:** *chi square, youtube, klasifikasi komentar youtube, algoritma naive bayes*

## ABSTRACT

*One of them is technology in the field of communication is a social media platform. Social media Youtube is one of the most widely used social media in Indonesia. The benefits gained by Content Creators or Youtubers come from AdSense. Youtube has several features provided such as likes, dislikes, views and comments (comments with negative or positive sentiments). An automatic classification system for YouTube comment sentiment is needed to classify positive comments and negative comments, while analyzing features that affect the number of subscribers so that Content Creators can find out features that can affect the number of subscribers. In this research a comment sentiment classification system will automatically be created using the Naive Bayes (NB) algorithm so that the process of classifying positive and negative comments can be done easily, the data used in the analysis are 53 Youtube channels with vlog video types. In addition, the data used as classification training data were 4166 positive sentiments and 4166 negative sentiments, after which an analysis of features affecting the number of subscribers was performed using chi square. The results of the analysis with chi square found there are 4 features that have an influence on the number of subscribers, namely the number of views with a chi square value of 23,105, dislike with a chi square value of 13,745, the number of positive sentiment comments with a chi square value of 18,123 and the number of likes with a chi square value of 13,745. The accuracy of the automatic classification system using Naive Bayes (NB) is 81%.*

**Keywords:** *chi square, youtube, youtube comment classification, naive bayes algorithm.*

## 1. PENDAHULUAN

*Content Creator* merupakan sebuah pekerjaan di dunia yang semakin modern ini, dengan sebuah video yang di *upload* pada media sosial Youtube, seseorang sudah bisa mendapatkan keuntungan. Keuntungan yang didapatkan tergantung dari jumlah *subscribers* dan *viewers* yang melihat video tersebut. *Content* yang dibuat sebaiknya *content* yang memiliki respon yang banyak sehingga jumlah *subscribers* dan *viewers* meningkat. Youtube memiliki fitur yang memungkinkan penonton dapat melakukan interaksi dengan *Content Creator* seperti, *like, dislike, view*

dan komentar yang memungkinkan penonton untuk memberikan komentar dengan sentimen positif atau negatif. Fitur-fitur yang memiliki pengaruh dapat dimanfaatkan untuk meningkatkan jumlah *subscribers* sehingga diperlukan analisis untuk mengetahui fitur-fitur yang dapat mempengaruhi jumlah *subscribers* dan sistem klasifikasi sentimen otomatis agar proses klasifikasi komentar jadi lebih mudah.

Penelitian sebelumnya yang menggunakan objek penelitian komentar Youtube dengan menggunakan *Text mining* dan algoritma Naive Bayes (NB) telah diterapkan (Luqyana, 2018) dan

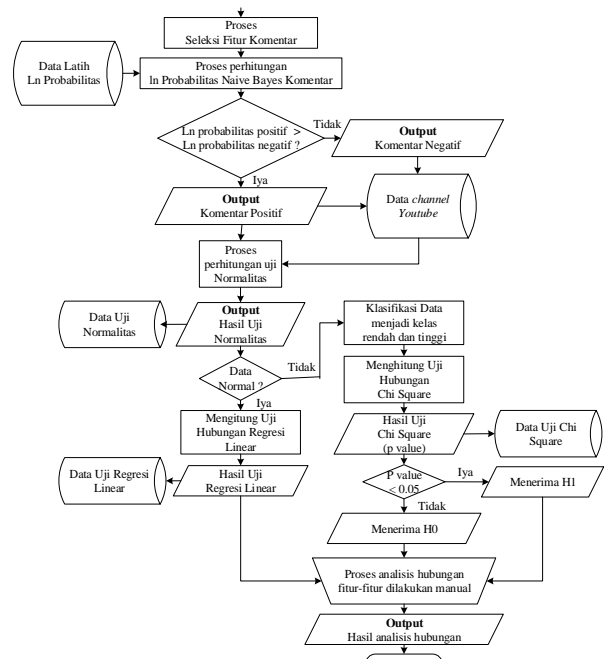
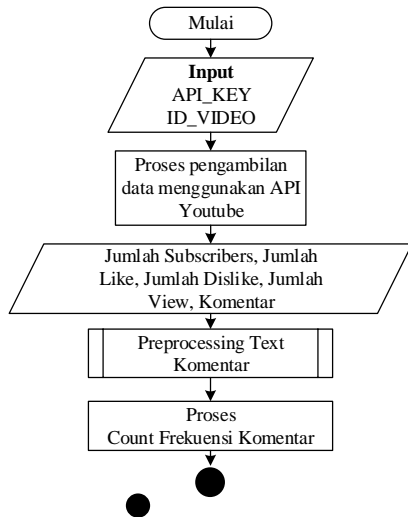
pernah dilakukan penelitian untuk dilakukan analisis komentar yang tujuannya mendukung strategi kampanye politik oleh (Evans, 2016). Namun belum ada penelitian yang membahas tentang preferensi yang dapat digunakan oleh *Content Creator* untuk mengetahui fitur-fitur yang dapat mempengaruhi jumlah *subscribers* dengan melihat jumlah *like*, jumlah *dislike*, jumlah *view*, jumlah komentar positif dan jumlah komentar negatif. Oleh karena itu pada penelitian ini akan dibuat rancang bangun sistem klasifikasi sentimen komentar Youtube otomatis, setelah itu akan dilakukan analisis fitur-fitur yang mempengaruhi jumlah *subscribers* menggunakan *chi square*.

Pada artikel ini akan dimulai dengan pendahuluan dan kajian dengan konsep dasar dari metode yang telah digunakan. Kemudian desain penelitian yang digunakan, hasil dan yang terakhir adalah kesimpulan.

## 2. METODOLOGI PENELITIAN

### Desain Sistem

Desain sistem dari penelitian ini dapat dilihat pada gambar 1. Desain diawali dengan pengumpulan data, proses *text mining*, klasifikasi sentimen menggunakan algoritma *Naive Bayes Classifier*, dan yang terakhir adalah dilakukan uji hubungan menggunakan *chi square*.



Gambar 1. Desain Sistem

Memuat rancangan penelitian atau desain penelitian, sasaran dan target penelitian (populasi dan sampel), teknik pengumpulan data dan teknik analisis data.

### Pengumpulan Data

Pada penelitian ini data yang dikumpulkan adalah jumlah *view*, jumlah *like*, jumlah *dislike*, jumlah komentar positif, jumlah komentar negatif dan jumlah *subscribers* untuk dapat dilakukan analisis hubungan. Agar komentar positif dan negatif diklasifikasi dengan otomatis maka diperlukan data latih komentar positif dan komentar negatif untuk dipakai sebagai data latih dalam algoritma *Naive Bayes Classifier*. Total jumlah keseluruhan komentar data latih dari 40 sampel/channel Youtube yang diambil adalah sebanyak 90.481 komentar. Kemudian data tersebut diklasifikasi secara manual untuk digunakan sebagai data latih.

Cara melakukan klasifikasi manual komentar positif dan negatif adalah dengan melihat isi dari komentar tersebut, apabila komentar tersebut berisi dukungan, pujian, ataupun saran agar *channel* tersebut berkembang maka komentar tersebut diklasifikasikan sebagai komentar positif, sebaliknya apabila komentar tersebut berisi penghinaan secara fisik ataupun saran yang tidak mendukung perkembangan *channel* maka komentar tersebut diklasifikasi sebagai komentar negatif. Berikut contoh cuplikan dilakukan klasifikasi positif atau negatif, dapat dilihat pada Tabel 1.

**Tabel 1.** Contoh Cuplikan Klasifikasi Komentar Manual

Komentar	Klasifikasi
“ Durasinya dipanjang lagi dong kak,”	Positif
“ Ah jelek lu.”	Negatif
“ Sukses terus yaa ayahanda”	Positif
“ Konten apa sih ini ga mendidik”	Negatif

Hasil dari klasifikasi manual didapatkan 50.217 komentar positif, 5.216 komentar negatif dan sisanya sentimen netral. Jumlah data uji yang digunakan adalah sebanyak 13 sampel/channel Youtube dengan jumlah komentar sebanyak 3.855.

Jumlah data yang digunakan sebagai data latih dan data uji dari 40 channel Youtube adalah 5.216 sentimen positif komentar Youtube dan 5.216 sentimen negatif komentar Youtube, setelah itu sentimen akan dibagi menjadi 2 bagian, yaitu 80% data latih dan 20% data uji dari 40 channel Youtube, data latih dan data uji dibuat seimbang agar data tidak cenderung mengarah kepada sentimen positif atau negatif, sedangkan jumlah data uji dari 13 channel Youtube adalah 3.855 sentimen.

### 3. HASIL DAN PEMBAHASAN

#### Perangkat Lunak

Aplikasi dapat melakukan analisis sentimen Youtube secara otomatis seperti terlihat pada gambar 2 dan gambar 3, setelah itu dapat melakukan perhitungan uji hubungan dengan menggunakan *chi square* seperti gambar 4. Inputan yang dimasukan adalah url video youtube yang akan dilakukan analisis, *quantity* adalah jumlah komentar yang akan diambil, komentar dapat diambil secara keseluruhan dengan mencentang *checkbox* “all”. Aplikasi dapat menambah data variabel jika *checkbox add* sampel dicentang “yes”. Setelah itu klik *analyze* untuk mulai melakukan analisis sentimen secara otomatis dan uji hubungan menggunakan *chi square*.

**Gambar 2.** Inputan Analisis

No	Comment	Ln Positive	Ln Negative	Sentiment	Validation
1	tuhan yesus sayang nang dengar lagu ikhlas maaf baik	-64.261	-83.782	POSITIVE	<input type="checkbox"/>

DATA YOUTUBE				
1	2983	0	21	66
Count Comment	Count View	Count Dislike	Count Like	Subscribers

**Gambar 3.** Hasil Klasifikasi Sentimen Otomatis

DATA CHI SQUARE				
24.000	14.519	14.519	18.963	2.667
VIEW SUBSCRIBERS	LIKE SUBSCRIBERS	DISLIKE SUBSCRIBERS	POSITIVE SUBSCRIBERS	NEGATIVE SUBSCRIBERS
Accept H <sub>0</sub>	Accept H <sub>0</sub>	Accept H <sub>0</sub>	Accept H <sub>0</sub>	Accept H <sub>0</sub>
There is a significant relationship between the number of dislikes and the number of subscribers	There is a significant relationship between the number of dislikes and the number of subscribers	There is a significant relationship between the number of dislikes and the number of subscribers	There is a significant relationship between the number of dislikes and the number of subscribers	There is no significant relationship between the number of dislikes and the number of subscribers

**Gambar 4.** Uji Chi Square

#### Uji Coba

Uji akurasi klasifikasi dilakukan dengan menggunakan 2,003 sentimen komentar data latih (40 sampel/channel Youtube). Dari hasil klasifikasi 2,003 sentimen, didapatkan 1,222 komentar diklasifikasi dengan benar sedangkan sisanya tidak tepat. Sedangkan uji akurasi klasifikasi dilakukan dengan menggunakan 3,855 sentimen komentar data uji (13 sampel/channel Youtube). Dari hasil klasifikasi 3,855 sentimen, didapatkan 3,578 komentar diklasifikasi dengan benar sedangkan sisanya tidak tepat. Jumlah akurasi yang didapatkan menurut perhitungan adalah sebagai berikut.

$$Akurasi\ 40\ Sampel = \left(\frac{1222}{2003}\right) \times 100\% = 61\%$$

$$Akurasi\ 13\ Sampel = \left(\frac{3578}{3855}\right) \times 100\% = 97\%$$

**Gambar 5.** Jumlah akurasi berdsarkan perhitungan

Uji akurasi klasifikasi dilakukan dua kali dikarenakan akurasi pada percobaan pertama atau pada data latih kurang akurat yaitu hanya 61%, sehingga ditambahkan 13 sampel baru atau pada data uji dan didapatkan akurasi yaitu 97%. Kemudian dilakukan perhitungan rata-rata akurasi yang didapatkan pada data latih dan data uji, sehingga didapatkan hasil akurasi sebagai berikut.

$$Total\ Rata - Rata\ Akurasi = \left(\frac{4800}{5855}\right) \times 100\% = 81\%$$

**Gambar 6.** Hasil akurasi

### 4. KESIMPULAN DAN SARAN

Ada 4 fitur yang mempengaruhi jumlah *subscribers* dan 1 Fitur yang tidak memiliki hubungan, yaitu.

#### Fitur View

Jumlah *view* dengan nilai *chi square* 23.105 dan nilai *p value*  $0.00 < 0.05$ , dapat disimpulkan bahwa ada hubungan antara jumlah *view* dan jumlah *subscribers*. Semakin tinggi jumlah *view* maka semakin tinggi juga jumlah *subscribers*, sebaliknya semakin rendah jumlah *view* maka semakin rendah juga jumlah *subscribers*. Menurut penelitian yang dilakukan oleh Hartanto Setiabudi (2018), motif dari orang-orang yang menonton atau orang-orang yang

menjadi *viewer content* Youtube tersebut adalah orang-orang yang kemungkinan besar memiliki motif untuk mendapatkan informasi dan ingin mendapatkan hiburan. Dapat disimpulkan jika *content* tersebut informatif dan menghibur, maka jumlah *view* akan semakin meningkat sehingga jumlah *subscribers* dapat meningkat.

#### **Fitur Like**

Jumlah *like* dengan nilai *chi square* 13.745 dan nilai *p value*  $0.00 < 0.05$ , dapat disimpulkan bahwa ada hubungan antara jumlah *like* dan jumlah *subscribers*. Semakin tinggi jumlah *like* maka semakin tinggi jumlah *subscribers*, sebaliknya semakin rendah jumlah *like* maka semakin rendah jumlah *subscribers*. Jumlah *like* yang tinggi dapat diartikan bahwa *content* tersebut disukai. *Content* Youtube banyak disukai kemungkinan besar dapat meningkatkan jumlah *viewer*, sehingga jumlah *subscribers* juga dapat meningkat.

#### **Fitur Dislike**

Jumlah *dislike* dengan nilai *chi square* 13.745 dan nilai *p value*  $0.00 < 0.05$ , dapat disimpulkan bahwa ada hubungan antara jumlah *dislike* dan jumlah *subscribers*. Semakin tinggi jumlah *dislike* maka semakin tinggi jumlah *subscribers*, sebaliknya semakin rendah jumlah *dislike* maka semakin rendah jumlah *subscribers*. Dari sampel yang terkumpul, setelah diamati, 21 *content* Youtube yang memiliki jumlah *dislike* tinggi dan jumlah *subscribers* tinggi ternyata memiliki jumlah *like* yang lebih tinggi daripada jumlah *dislike*. Sehingga dapat diartikan bahwa *content* tersebut masih bersifat positif (informatif dan menghibur). Sehingga dapat disimpulkan, jika jumlah *dislike* tidak lebih tinggi dari jumlah *like* (jumlah *like* harus lebih tinggi) maka kemungkinan besar dapat meningkatkan jumlah *subscribers*.

#### **Fitur Komentar Positif**

Jumlah komentar positif dengan nilai *chi square* 18.123 dan nilai *p value*  $0.00 < 0.05$ , dapat disimpulkan bahwa ada hubungan antara jumlah komentar positif dan jumlah *subscribers*. Semakin tinggi jumlah komentar positif maka semakin tinggi jumlah *subscribers*, sebaliknya semakin rendah jumlah komentar positif maka semakin rendah jumlah *subscribers*. Komentar positif dapat membuktikan bahwa *content* tersebut informatif dan menghibur. Fitur ini hampir sama dengan fitur *like*, namun fitur ini diungkapkan dalam bentuk *text*. Sehingga dapat disimpulkan bahwa *content* Youtube

yang bersifat positif (informatif dan menghibur) dapat meningkatkan jumlah *subscribers*.

#### **Fitur Komentar Negatif**

Jumlah komentar negatif dengan nilai *chi square* 2.277 dan nilai *p value*  $0.131 > 0.05$ , dapat disimpulkan tidak ada hubungan antara jumlah komentar negatif terhadap jumlah *subscribers*. Hal tersebut mungkin terjadi karena penonton lebih berhati-hati dalam memberikan komentar negatif dan lebih memilih untuk melakukan *dislike*. Indonesia sendiri telah mengeluarkan Undang-Undang tentang *cyber bullying*, pada pasal 27 ayat (3) UU ITE (Informasi Transaksi Elektronik) menyebutkan melarang setiap orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan dan/atau membuat dapat diaksesnya informasi elektronik dan/atau dokumen yang memiliki muatan penghinaan dan/atau pencemaran nama baik. Jumlah komentar negatif tidak berpengaruh karena penonton lebih berhati-hati untuk memberikan komentar negatif sehingga lebih memilih untuk melakukan *dislike*.

## **5. REFERENSI**

- [1] EVANS, W (2016). Analisis Konten Pada Media Sosial Video Youtube Untuk Mendukung Strategi Kampanye Politik Dengan Menggunakan Analisis Sentimen
- [2] FEBRUARIYANTI, H., & ZULIARSO, E. (2012). Klasifikasi dokumen berita teks bahasa indonesia menggunakan ontologi. *Dinamik*, 17(1).
- [3] GO, A., HUANG, L., & BHAYANI, R. (2009). Twitter Sentiment Analysis. Final Project Report, Stanford University, Department of Computer Science.
- [4] KONCZ, P., & PARALIC, J. (2011). An approach to feature selection for sentiment analysis. In 2011 15th IEEE International Conference on Intelligent Engineering Systems (pp. 357–362).
- [5] LIANTONI, F., NUGROHO, H. (2015). Klasifikasi Daun Herbal Menggunakan Metode Naive Bayes Classifier Dan Knearest.
- [6] LESMANA, P. I. (2013). Analisis Sentimen Pengguna Layanan Media Sosial Twitter Di Indonesia.
- [7] MONARIZQA, N., NUGROHO, L. E., & HANTONO, B. S. (2014). Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro dan Teknologi Informasi*, 1(3).

- [8] NEIGHBOR. LUQYANA, W. A., CHOLISSODIN, I., PERDANA, R. S., (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine.
- [9] PRASETYO, E. (2012), Data Mining konsep dan Aplikasi menggunakan Matlab.
- [10] RETNAWIYATI, E., M.KOM, F. M., & NEGARA, M.KOM, E. S. (2016). Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk.
- [11] VIJAYARANI, S., & DHAYANAND, S. (2015). Data Mining Classification Algorithms for Kidney Disease Prediction.
- [12] International Journal on Cybernetics & Informatics (IJCI), 4(4), 13-25.
- [13] HARTANTO S. (2018), Motif Subscriber Menonton Channel Youtube (Studi Deskriptif Kualitatif Motif Subscriber Menonton Channel Youtube Presiden Jokowi).