

## ANALISIS SENTIMEN TWITTER UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (NBC) PADA SENTIMEN R PROGRAMMING

Susanti Fransiska<sup>1)</sup>, Yolanda<sup>2)</sup>

Program Studi Informatika Fakultas Teknik Universitas Siliwangi  
Jln. Siliwangi No. 24 Kahuripan, Tasikmalaya  
email: 167006031@student.unsil.ac.id<sup>1)</sup>, 167006021@student.unsil.ac.id<sup>2)</sup>

### Abstrak

Twitter merupakan salah satu media sosial yang memungkinkan penggunaanya bebas berkomentar atau menulis *tweet* opini tentang apapun yang terjadi, termasuk opini tentang film. Banyaknya *tweet* komentar atau opini yang ditulis di twitter dapat dimanfaatkan untuk mendapatkan informasi dan penilaian film yang akan menjadi referensi bagi pengguna twitter lain. Akan tetapi dengan banyaknya *tweet* membuat kesulitan bagi pengguna yang ingin melihat opini tentang review film tersebut karena data *tweet* masih terusun secara acak. Agar diketahui opini secara tersusun sesuai dengan klasifikasi maka dalam penelitian ini dilakukan analisis sentimen dengan metode Naive Bayes. Judul film yang menjadi fokus dalam penelitian ini adalah "Parasite". Data yang dihasilkan dari tahap analisis sentimen ini dikelompokkan menjadi dua jenis sentimen. Pertama *sentiment emotion* dengan mengklasifikasikan data menjadi enam: *emotion, anger, disgust, pear, joy, sadness, surprise*. Kedua *sentiment polarity* mengklasifikasikan data menjadi sentimen positif, sentimen negatif dan sentimen netral. Dengan dua klasifikasi yang berbeda didapatkan hasil data dari *sentiment emotion* lebih banyak *unknown* dan untuk *sentiment polarity* lebih banyak sentimen negatif. Ukuran data dari data awal yang sebanyak 99.521 *tweet* dengan ukuran 13.9Mb menjadi 7.5MB dan data *tweet* menjadi sebanyak 36.016.

**Kata Kunci :** Analisis Sentimen, *Naive Bayes Classifier*, Twitter, R.

### Abstract

Twitter is a social media that allows users to freely comment or write opinion tweets about whatever happens, including opinions about films. The number of comments or opinions tweets written on Twitter can be used to get information and movie ratings that will be a reference for other Twitter users. However, the number of tweets makes it difficult for users who want to see opinions about the film review because the tweet data is still arranged randomly. In order to know the opinion arranged in accordance with the classification, in this study sentiment analysis using the Naive Bayes method. The title of the film that is the focus of this research is "Parasite". Data generated from the sentiment analysis stage are grouped into two types of sentiments. First emotion sentiment by classifying data into six: *emotion, anger, disgust, pear, joy, sadness, surprise*. Both polarity sentiments classify data into positive sentiments, negative sentiments and neutral sentiments. With two different classifications obtained data from emotion sentiments are more unknown and for sentiment polarity more negative sentiments. Data size from the initial data were 99,521 tweets with a size of 13.9MB to 7.5MB and tweet data to be 36,016.

**Keywords:** Sentiment Analysis, *Naive Bayes Classifier*, Twitter, R.

### 1. PENDAHULUAN

Pada masa ini masyarakat modern sangat bergantung pada penggunaan teknologi digital, memiliki dan menggunakan media sosial. Media sosial merupakan sebuah media *online* yang penggunaanya dapat dengan mudah berbagi dan menciptakan isi meliputi blog, jejaring sosial, wiki forum dan virtual. Salah satu media sosial yang banyak digunakan yaitu Twitter.

Twitter merupakan salah satu media sosial *microblog* yang memungkinkan penggunaanya untuk

menulis tentang berbagai topik dan membahas isu-isu yang sedang terjadi. Twitter memiliki banyak pengguna aktif sehingga akan memberikan komentar atau informasi terbaru tentang hal yang sedang diperbincangkan di dunia dan menimbulkan *trending topic twitter*.

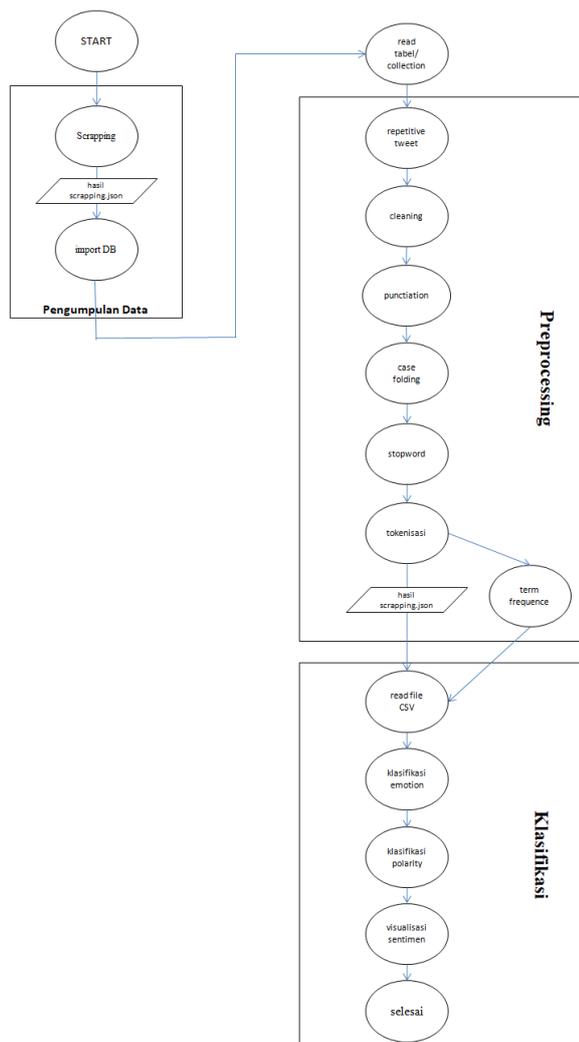
Tingginya pengguna twitter menjadikan peluang untuk membagikan informasi, ide gagasan, menuangkan isi hati atau tentang hiburan termasuk film. Pengguna dapat memberikan komentar atau opini tentang film yang sedang tayang di bioskop dan memberikan penilaian akan film tersebut. Informasi

tweet tersebut akan menjadi referensi bagi pengguna twitter lain. Akan tetapi dengan banyaknya tweet yang banyak dan tersusun secara acak menyebabkan kesulitan bagi pengguna twitter untuk mengetahui opini positif, negatif dan netral.

Banyaknya pengguna Twitter yang menyampaikan komentar atau opini tersebut dapat dimanfaatkan dengan cara mengklasifikasikan review tweet sehingga diperoleh sebuah informasi. Salah satu metode yang dapat digunakan untuk menganalisis opini-opini di tweet adalah analisis sentimen. Data yang didapat dari twitt diproses dengan *text mining*, dilanjutkan dengan mengklasifikasikan data tweet ke dalam tiga kelas, positif, negatif dan netral. Dalam penelitian ini pengklasifikasian menggunakan algoritma Naive Bayes Classifier (NBC).

## 2. METODOLOGI

Tahapan yang dilakukan dalam penelitian ini ditampilkan pada gambar 1.



Gambar 1. Metodologi Analisis Sentiment

### 2.1 Pengumpulan Data

- Pengumpulan Data Twitter (*Scraping*)  
Data tweet yang diperoleh adalah data yang secara langsung diambil dari situs twitter dengan keyword “parasitemovie” atau “parasite” menggunakan *twitterscrapper python*.
- Menyimpan data ke *NoSQL MongoDB*.  
Data hasil *scraping* diimport ke *database* dengan bantuan *tools Studio 3t*.

### 2.2 Preprocessing

Ada beberapa tahapan dalam *preprocessing*, diantaranya:

- Penghapusan Data Berulang (*Repetitive Tweet*).
- Pembersihan Data (*Cleaning*), seperti penghapusan *url*, *new line*, dan karakter-karakter lain seperti *amp*.
- Penghapusan *Punctuation* (tanda baca).
- Case Folding*.
- Penghapusan *stopword* (*slang word*).
- Tokenisasi.
- Term *frequency* (kata yang sering muncul) dan visualisasinya.

### 2.3 Klasifikasi dan Pelabelan

Klasifikasi dan pelabelan data menggunakan *function* yang tersedia pada *package* ‘sentimen’ *R programming* dengan parameter *algorithm bayes*.

### 2.4 Visualisasi Sentiment

Visualisasi menggunakan *package* *wordcloud* dan *ggplot2* pada *R programming*.

## 3. HASIL DAN PEMBAHASAN

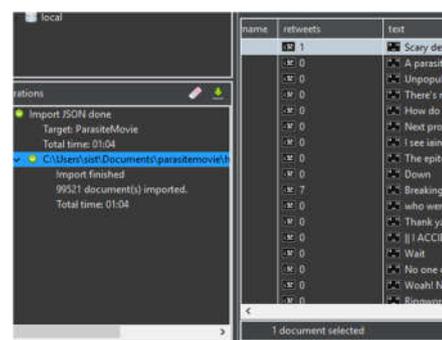
### 3.1 Hasil Scraping

Proses *Scraping* mendapatkan 99.521 data *tweet* dalam format *file JSON* dengan ukuran 103 MB seperti ditampilkan pada gambar 2.



Gambar 2. File Hasil Scraping

Data hasil *scraping* diimport ke *Database NoSQL MongoDB* dengan waktu 1.4 menit seperti ditampilkan pada gambar 3.



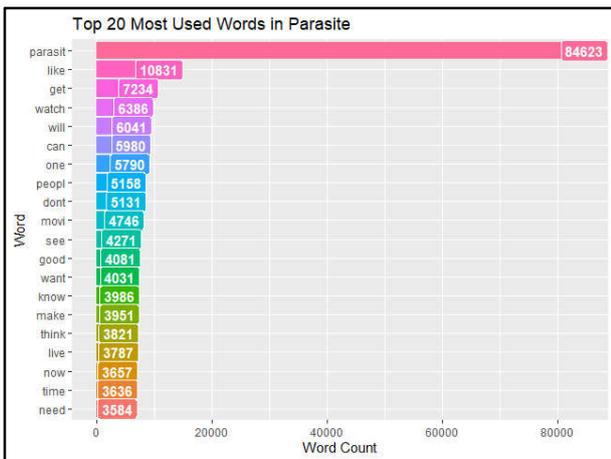
Gambar 3. Import data ke database

3.2 Proses Analisis

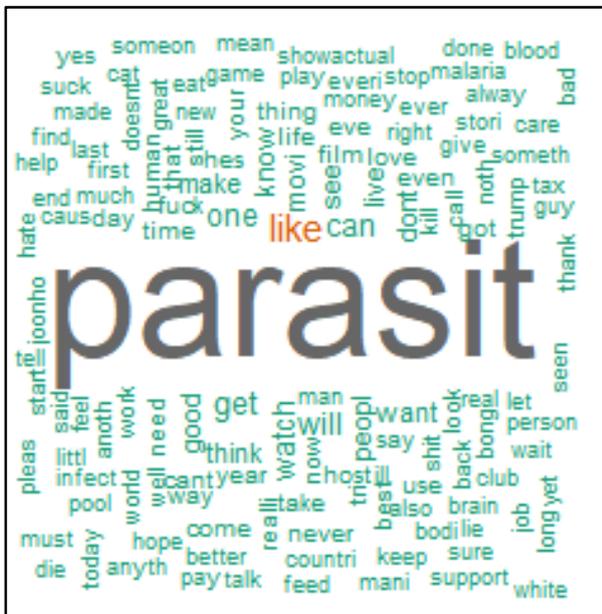
Pada tahap *preprocessing* hanya data dalam kolom 'text' yang dipilih dari *Database MongoDB* dengan ukuran data 13.9Mb. Setelah melakukan tahap *preprocessing* ukuran data menjadi 7.5MB dan data *tweet* menjadi sebanyak 36.016 dari sebelumnya 99.521 seperti ditampilkan pada gambar 4.

texttweet	99521 obs. of 1 variable
top20	20 obs. of 2 variables
twitclean	Large SimpleCorpus (7.5 Mb)
uncleanedTweets	Large SimpleCorpus (13.9 Mb)
unique_tweet	96016 obs. of 1 variable

Gambar 4. Hasil Proses Preprocessing



Gambar 5. Visualisasi hasil term frequency



Gambar 6. Wordcloud hasil preprocessing

3.3 Hasil Analisis

Analisis sentimen yang dilakukan menghasilkan *sentiment emotion* dan *polarity*. *Sentiment emotion* mengklasifikasikan data menjadi enam *emotion*, *anger*, *disgust*, *pear*, *joy*, *sadness*, *suprise* seperti ditampilkan pada gambar 7.

	ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE	BEST_FIT
1	1.46871776464786	3.09234031207392	7.34083555412328	1.02547755260094	1.7277074477352	2.78695866252273	fear
2	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	joy
3	1.46871776464786	3.09234031207392	2.06783599555953	13.6561935556456	7.34083555412328	2.78695866252273	joy
4	1.46871776464786	3.09234031207392	2.06783599555953	7.34083555412328	1.7277074477352	2.78695866252273	joy
5	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	joy
6	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	joy
7	1.46871776464786	3.09234031207392	2.06783599555953	7.34083555412328	1.7277074477352	2.78695866252273	joy
8	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	15.5670917660994	2.78695866252273	sadness
9	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	joy
10	1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	joy
11	1.46871776464786	3.09234031207392	2.06783599555953	7.34083555412328	1.7277074477352	2.78695866252273	joy

Gambar 7. Hasil perhitungan klasifikasi *sentiment (emotion)* dengan algoritma naive bayes

Analisis *polarity* mengklasifikasikan data menjadi *positive* dan *negative* seperti ditampilkan pada gambar 8.

	POS	NEG	POS/NEG	BEST_FIT
1	16.5333679773693	26.1492093145274	0.632270283147112	negative
2	1.03127774142571	34.4860789518123	0.029904175040216	negative
3	24.9775602759011	53.2392597680621	0.469156791148418	negative
4	9.47547003995745	8.78232285939751	1.07892526745566	neutral
5	1.03127774142571	0.445453222112551	2.31512017476245	positive
6	1.03127774142571	8.78232285939751	0.117426534862834	negative
7	17.9196623364692	17.8123396772424	1.00602518608961	neutral

Gambar 8. Hasil perhitungan klasifikasi *sentiment (polaritas)* dengan algoritma naive bayes

Hasil klasifikasi dan pelabelan *polarity* dan *emotion* pada setiap *tweet* ditampilkan pada gambar 9.

	text.X	text.text	emotion	polarity
1	1001	marriage mostly draining pick right parasite	joy	negative
2	1002	symptoms think parasite im freaking much	unknown	neutral
3	1003	one little freaks got a kind disease parasite im really pl...	joy	negative
4	1004	yeah like see people trying rape girl hes rapist like ok ...	joy	negative
5	1005	dnc hillary clinton paid russia fake dossier barry nothing ...	joy	negative
6	1006	dua garis biru parasite spiderman far home	unknown	negative
7	1007	u cant stand anything zombie u parasite winning car...	joy	neutral
8	1008	leishmania tarentolae taxonomic classification application ...	unknown	positive
9	1009	cyclospora news lately need know parasite dr todd el...	unknown	negative
10	1010	hypercore parasite world tra underground distorted	unknown	neutral
11	1011	youre telling post scorer wont work paint sharp youre ...	joy	negative
12	1012	destroying something hasnt bom world yet actually par...	joy	negative

Gambar 2. Hasil Klasifikasi dan Pelabelan Sentimen

