

# Perbandingan Algoritma Word Matching dan Naive Bayes untuk Klasifikasi Sentimen Analisis Komentar Instagram

Agnes Setyawinda<sup>1\*</sup>, Barep Setiyadi<sup>2</sup>, Anggit Dwi Hartanto<sup>3</sup>

<sup>1,2,3</sup> Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta

Email: <sup>1</sup>agnes.3843@students.amikom.ac.id, <sup>2</sup>barep.s@students.amikom.ac.id, <sup>3</sup>anggit@amikom.ac.id

**Abstrak** – Analisis sentimen telah menunjukkan bahwa otomatisasi dan pengenalan komputasi terhadap sentimen adalah mungkin dan berkembang seiring berjalannya waktu, karena faktor munculnya tren teknologi baru dan keadaan yang semakin dinamis dari bahasa manusia sebagai bentuk komunikasi. Dengan adanya media sosial semakin banyak pula teks-teks berupa data informal, menyebabkan proses ekstraksi dan penguraian informasi yang relevan menjadi masalah. Oleh karena itu pada penelitian ini penulis mengusulkan dua metode klasifikasi yang kemudian akan melakukan perbandingan hasil dari kedua metode tersebut.

**Kata Kunci:** Analisis Sentimen, media sosial, R, Informasi .

## I. PENDAHULUAN

Hootsuite mencatat pengguna sosial media aktif di Indonesia pada tahun 2019 tercatat 150 Juta pengguna artinya mendominasi penetrasi sebanyak 56% penduduk Indonesia, melalui postingan atau komentar untuk mengutarakan pendapat dari berbagai sudut pandang dan subjek tertentu. Dari berbagai komentar yang terkumpul bisa memberikan tinjauan singkat pandangan warganet terhadap suatu subjek di media sosial. Seperti pendapat masyarakat terhadap suatu produk atau tokoh publik seberapa baik sentimen yang diberikan oleh warganet terhadap subjek tersebut. Data yang dikumpulkan dari komentar warganet terhadap suatu subjek tertentu, sumber data ini dapat dimanfaatkan lebih lanjut dengan metode sentimen analisis. Menurut KBBI [1], Sentimen merupakan /sen·ti·men/ /séntimén/ pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu. Berdasarkan Kamus Oxford [2], Analisis sentimen adalah prosedur membedakan dan mengklasifikasikan secara komputasi perasaan dikomunikasikan dalam sedikit konten, terutama dengan tujuan akhir untuk memutuskan apakah disposisi penulis terhadap tema, subjek tertentu tergolong bersifat netral, positif atau negatif. Penelitian mengenai analisis sentimen sudah dilakukan pada penelitian-penelitian sebelumnya seperti Pang dan Lee [3], Narad Hipa et al. [4], dan Gamalo et al [5]. Baru-baru ini, penelitian mengenai analisis

sentimen berfokus pada klasifikasi konten yang merujuk pada suatu subjek. Saat ini dalam aplikasi bisnis, timbal balik pengguna berupa komentar atau dikenal sebagai sentimen telah digunakan dalam berbagai bidang seperti politik, fasilitas publik, tokoh publik, serta layanan yang berupa jasa atau produk. Emosi pengguna dan perasaan dalam komentar digunakan untuk menciptakan pengetahuan yang dapat digunakan untuk menganalisis timbal balik publik terhadap tokoh publik, fasilitas publik atau suatu merek tertentu. Adapun penelitian tentang metode teknik klasifikasi dengan memperhatikan setiap kelebihan dan kekurangan sebagai pertimbangan untuk memilih teknik klasifikasi dalam pengolahan data yang baik [6], dan terbukti dalam penelitiannya bahwa menggunakan metode klasifikasi naive bayes menghasilkan nilai akurasi terbaik [7][8]. Selain menggunakan beberapa algoritma klasifikasi seperti naive bayes ada metode lain untuk melakukan klasifikasi sentimen sebuah kata yaitu dengan menggunakan metode word matching atau pencocokan kata secara sederhana. Dalam penelitian ini penulis akan membandingkan klasifikasi dengan menggunakan metode *word matching* dengan bantuan opinion word lexicon dengan algoritma naive bayes disini peneliti menggunakan pustaka sentimen pada bahasa R, dimana data komentar yang sudah diambil akan diproses dan di konversi kedalam format yang lebih terstruktur terlebih dahulu, menghilangkan kata berulang, kata sambung dan tanda baca [9], baru kemudian bisa dilakukan klasifikasi emosi dan sentimen setiap kata.

## II. PENELITIAN YANG TERKAIT

Tahapan *preprocessing* atau *data processing* adalah langkah penting dalam melakukan klasifikasi analisis sentimen karena langkah ini akan menentukan format data dan kata yang mempengaruhi klasifikasi. Langkah *data processing* untuk menghilangkan teks yang tidak diperlukan oleh peneliti sebelumnya[7] menggunakan model ejaan untuk mengoreksi teks sesuai dengan format yang benar dan memperbaiki ejaan yang salah atau kata yang salah ketik. Naradhipa dan Purwarianti[4] melakukan pembersihan data numerik, konversi dari numerik ke teks, menghapus duplikasi teks.

\*) penulis korespondensi: Agnes Setyawinda  
Email: agnes.3843@students.amikom.ac.id

Lunando dan Purwarianti[8] memanfaatkan konversi karakter numerik ke dalam alfabetik, menghapus kata berulang, dan konversi dari kata informal ke dalam kata formal. Gamallo et al. [5] menggunakan metode Naive Bayes untuk mengklasifikasikan sentimen twitter dengan akurasi 67% . Secara umum, klasifikasi sentimen menggunakan *machine learning*. Naive bayes adalah metode yang paling banyak digunakan untuk klasifikasi teks.

### III. METODE PENELITIAN

Langkah langkah penelitian untuk melakukan sentimen pertama adalah pengambilan data, preprocessing, termasuk ekstraksi data dan klasifikasi. Langkah tersebut dapat dilihat pada Gambar 1.



**Gambar 1.** Langkah Klasifikasi

#### A. Instagram Data Scrapping

Pada penelitian ini, peneliti menggunakan *instagram-scrapper* untuk melakukan pengambilan data yang relevan dengan topik yang akan diambil. Pada kasus ini topik yang diambil adalah akun resmi dari Tirtol ID. Data yang diambil adalah komentar pada Tirtol ID dari tanggal 14 Juli 2019 sampai dengan 9 Agustus 2019 total data keseluruhan data yang diambil 22.0071 terdiri dari *id*, *post\_time*, *is\_video*, *caption*, *tags*, *n\_tags*, *video\_view*, *media\_like*, *comments\_username*, *comments\_time*, *comments\_text*.

<b>Id</b>	1	1	1
<b>Post_time</b>	2019-08-09 19:06:05	2019-08-09 19:06:05	2019-08-09 19:06:05
<b>Is_video</b>	TRUE	TRUE	TRUE
<b>Caption</b>	Soal impor rektor, kamu pro atau kontra? #Tirtografi	Soal impor rektor, kamu pro atau kontra? #Tirtografi	Soal impor rektor, kamu pro atau kontra? #Tirtografi
<b>Tags</b>	Tirtografi	Tirtografi	Tirtografi
<b>N_Tags</b>	1	1	1
<b>Video_view</b>	7663	7663	7663
<b>Media_like</b>	913	913	913
<b>Comments_username</b>	fariszfirst	nadyaso	ikansaputro

Comments_time	2019-08-09 19:38:04	2019-08-09 19:38:11	2019-08-09 19:41:30
Comments_text	@bayuprasetyop wkwkwk ya bukan gitu maksudnya 😊	Kebijakan menteri ini menurut saya scr tersirat bahwa rektor dr negara sendiri kurang mampu membawa PTNnya menembus ranking 100 besar di dunia. Harusnya dianalisis apa yg menjadi penyebabnya ...	Orang lingkungan dosen di kampus tuh udah kayak militer, senioritas kentel banget, mau impor rektor ya si rektornya nnt dikerjain sama yg lebih lama dsna

Tabel. 1 Data Hasil *Scraping*.

#### B. Data Preprocessing

Sosial media memberikan kesempatan kepada seluruh pengguna untuk membagikan pendapatnya. Komentar pada *instagram* dari *instagram-scrapper* berbentuk berkas *JSON* yang tidak terstruktur. contohnya *post-time* yang masih terformat *epochtime*, penggunaan simbol emotikon, dialek lokal, duplikasi data, dan kesalahan pada penulisan komentar. ekstraksi data akan berperan untuk konversi dari berkas *JSON* kedalam bentuk teks terformat *csv* dimana berkas *csv* lebih terstruktur dibandingkan dengan *JSON*. dengan bantuan *R* dan beberapa pustaka seperti *jsonlite*, *listviewer*, dan *tidyverse*. Ekstraksi data melalui langkah-langkah berikut :

1) Data *Rectangling*  
Dalam penelitian ini peneliti menggunakan bantuan bahasa pemrograman *R* dan beberapa pustaka seperti *jsonlite*, *listviewer*, dan *tidyverse* untuk melakukan Data *rectangling* kedalam format tabulasi.

2) Ekstraksi Data  
Selanjutnya melakukan ekstraksi data menggunakan fungsi ‘map()’ dari pustaka ‘purr’. baru kemudian data hasil ekstraksi tersebut dapat disimpan sebagai kolom baru menggunakan fungsi ‘mutate()’ dari paket ‘dplyr’. Berikut hasil dari ekstraksi data yang kemudian disimpan kedalam kolom baru

Id	post	post_time	is_video	caption	tags	video_view	media_like
1	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [1]>	<int [1]>	<int [1]>
2	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [5]>	<int [1]>	<int [1]>
3	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [4]>	<int [1]>	<int [1]>
4	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [3]>	<int [1]>	<int [1]>
5	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [4]>	<int [1]>	<int [1]>
6	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [4]>	<int [1]>	<int [1]>
7	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [4]>	<int [1]>	<int [1]>
8	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [1]>	<int [1]>	<int [1]>
9	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [4]>	<int [1]>	<int [1]>
10	<list [22]>	<int [1]>	<ig [1]>	<chr [1]>	<chr [1]>	<int [1]>	<int [1]>

1-10 of 100 rows | 1-8 of 11 columns Previous [ 1 ] 2 3 4 5 6 ... 10 Next

**Gambar 2.** Hasil Ekstraksi Data

Data hasil ekstraksi diatas masih menunjukkan tipe data berupa 'list', tipe kolom tersebut harus diubah menjadi tipe yang sesuai dengan data yang ada, perlu dilakukan *flattening* pada hasil ekstraksi di atas menggunakan fungsi 'smart\_extract()' . data hasil ekstraksi seperti pada **Gambar 3**

Id	post	post_time	is_video
1	<list [22]>	1563352365	TRUE
2	<list [22]>	1563337665	FALSE
3	<list [22]>	1563312474	FALSE
4	<list [22]>	1563263163	TRUE
5	<list [22]>	1563263666	FALSE
6	<list [22]>	1563237436	FALSE
7	<list [22]>	1563179266	FALSE
8	<list [22]>	1563154064	FALSE
9	<list [22]>	1563139665	FALSE
10	<list [22]>	1563099305	TRUE

1-10 of 100 rows | 1-4 of 11 columns Previous 1 2 3 4 5 6 ... 10 Next

**Gambar 3.** Hasil *flattening*

### 3) Pembersihan Data

Data hasil ekstraksi perlu dilakukan pembersihan pada beberapa kolom, diantaranya *straightforward processing* pada kolom *post\_time* yaitu mengubah bentuk *epochtime* kedalam bentuk waktu yang terformat seperti pada **Gambar 4**, *lengthy processing* pada kolom *caption* dengan menghapus string berupa ("\\n") bertujuan untuk menghapus tanda '\n' dan spasi berlebih pada kolom *caption* menggunakan fungsi 'str\_trim()' seperti pada **Gambar 5**. Kemudian menghitung jumlah *tags* yang digunakan dan menyimpan sebagai kolom baru *n\_tags* seperti pada **Gambar 6**.

**Gambar 4.** Hasil konversi *epochtime*

post_time
2019-08-09 19:06:05
2019-08-09 15:01:05
2019-08-09 08:01:14
2019-08-08 21:06:03
2019-08-08 19:01:06
2019-08-08 11:10:36
2019-08-07 19:01:06
2019-08-07 12:01:04
2019-08-07 08:01:05
2019-08-06 20:48:25

**Gambar 4.** Hasil konversi *epochtime*

caption
Soal Impor rektor, kamu pro atau kontra? #TirtoGraf
Balik kekuasaan kekuasaan kekuasaan kekuasaan kekuasaan kekuasaan ...
Ada petir di balik lelucon para pemancing tawa. Karena menjadi lucu tak sama dengan menjadi baha... #infografik #humor...
• Denger-denger sekarang lo punya bisnis gitu, ya? Jujuan apa lo? Kopi? Cendol? Baju? ☺️ Data kependudukan. nhe~Nin...
Lansia di Kubu adalah lansia yang bahagia, semangat, dan tetap sehat meski sudah memasuki usia senja. Hal ini tentu men...
Udah mertua memang buat memberhiskan udara di dalam ruangan. Buat di luar ruangan juga bisa, sih. Tp, perlu b... Selain prestasi dan keberlebihannnya dalam hidup ini, ternyata bau badan juga bisa menentukan bagaimana kelihatan kelompok... Serat sungguh berguna buat melancarkan pencernaan. Dan, tubuh perlu sebanyak 20-35 gram serat setiap hari lho. Sudah ... Sering lihat, tapi engak sin meme ini diambil dari mana? Pam pam pam~ #infografik #adademie #sosialmedia #memes
Mail di lampu 🌟 adun 🌟 gelap-gelapan ➡ jidinya ✅ seperti 🌟 siluman 🌟 #TirtoGraf

1-10 of 100 rows | 4-4 of 11 columns Previous 1 2 3 4 5 6 ... 10 Next

**Gambar 5.** Hasil *Lengthy Processing*

tags	n_tags	video_view
TirtoGraf	1	7663
infografik, bokeh, pornografi, bisnis, ekonomi	5	NA
infografik, pelawak, humor, lucu	4	NA
TirtoGraf	1	12127
lanjutusia, muda, Lansia, foreveryoung	4	NA
infografik, tanaman, polusi, lidahmerutua	4	NA
infografik, baubadan, parasite, sosial	4	NA
ads	1	NA
infografik, meme, sosialmedia, deadmemes	4	NA
TirtoGraf	1	17594

**Gambar 6.** Konversi *tags* sebagai *string*

### 4) Konversi *JSON* ke dalam bentuk *CSV*

Konversi *JSON* ke dalam bentuk *CSV* menggunakan fungsi 'write\_csv()'

### C. Klasifikasi Sentimen

Data komentar pengguna yang sudah melalui tahap *preprocessing* akan diklasifikasikan kedalam kategori positif, negatif, netral dan dalam bentuk ekspresi emosi seperti marah, sedih, senang, enjoy.

### 1) Word Matching

Metode *Word Matching* dilakukan dengan menggunakan *opinion-lexicon* kamus kata positif dan negatif dalam hal ini menggunakan kamus bahasa Indonesia dari ID-OpinionWords[10][11]. metode *word matching* dilakukan dengan cara sederhana yaitu dengan mencocokan kata dengan kamus yang ada, kemudian dilakukan *scoring* secara sederhana, jika *score* sentimen lebih dari 0 (nol) maka dikatakan sebagai sentimen positif, jika *score* sentimen bernilai kurang dari 0 (nol) maka sentimen bernilai negatif, namun apabila sentimen *score* bernilai sama dengan 0 (nol) maka sentimen bernilai netral.

### 2) Naive Bayes Kategori Emosi Komentar

Untuk melakukan klasifikasi analisis sentimen Naive Bayes menggunakan R, membutuhkan pustaka Rstem dari Omegahat dan Sentiment dari CranR Project. Untuk mendapatkan hasil klasifikasi emosi menggunakan fungsi 'classify\_emotion()'. maka akan didapat hasil seperti pada **Gambar 7**.

ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE	BEST_FIT
1.46871776464786	3.09234031207392	2.06783599555953	7.34083555412328	1.7277074477352	2.78695866252273	joy
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	NA
7.34083555412328	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	anger
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	NA
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	NA
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	NA
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	anger
7.34083555412328	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	anger
1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	NA

**Gambar 7.** Klasifikasi emosi dengan Naive Bayes

### 3) Naive Bayes Klasifikasi Polaritas Komentar

Dengan memanfaatkan pustaka *sentiment* dan algoritma Naive bayes, klasifikasi polaritas komentar bisa dilakukan dengan menggunakan fungsi 'classify\_polarity()' untuk klasifikasi polaritas, komentar akan dibagi menjadi dua bagian yaitu sentimen positif dan negatif akan didapat hasil seperti pada **Gambar 8**.

POS	NEG	POS/NEG	BEST_FIT
1.03127774142571	0.445453222112551	2.31512017476245	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
8.78232285939751	0.445453222112551	19.7154772340574	positive
1.03127774142571	9.47547003995745	0.108836578774127	negative
8.78232285939751	0.445453222112551	19.7154772340574	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
8.78232285939751	0.445453222112551	19.7154772340574	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive
9.47547003995745	0.445453222112551	21.2715265477714	positive
1.03127774142571	0.445453222112551	2.31512017476245	positive

**Gambar 8.** Klasifikasi Polaritas Komentar

dengan melakukan klasifikasi polaritas komentar bisa diketahui kecondongan komentar kearah negatif, positif, atau netral.

### IV. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini diambil dari komentar instagram akun Tirto Id sebanyak 22.0071 komentar.

#### A. Word Matching

Metode Word Matching dilakukan dengan menggunakan opinion-lexicon kamus kata positif dan negatif dalam hal ini menggunakan kamus bahasa Indonesia dari ID-OpinionWords[10][11]. metode word matching dilakukan dengan cara sederhana yaitu dengan mencocokan kata dengan kamus yang ada, kemudian dilakukan scoring secara sederhana, jika score sentimen lebih dari 0 (nol) maka dikatakan sebagai sentimen positif, jika score sentimen bernilai kurang dari 0 (nol) maka sentimen bernilai negatif, namun apabila sentimen score bernilai sama dengan 0 (nol) maka sentimen bernilai netral.

Metode Word Matching dilakukan dengan mencocokan kamus data kata positif dan negatif menggunakan kamus ID-OpinionWords, metode ini cenderung lebih rumit karena dilakukan dengan pencocokan manual, hasil dari pencocokan seperti pada **Gambar 9**.

@bayuprasetyo wkwkwk ya bukan gitu maksudnya	1
Kebijakan menteri ini menurut saya scr terisrat bahwa rektor dr neg	1
Orang lingkungan dosen di kampus tuh udah kayak militer seniorita	1
Kenapa nggak ambil rektor orang Indo yang lulusan luar aja? Macar	1
Kontra... Gak terlalu penting sih untuk menaikkan ranking PTN... yg	-1
Pro, buat bahan acuan di masukin ke salah satu universitas yang k	-1
@farahmabilar vice versa	0

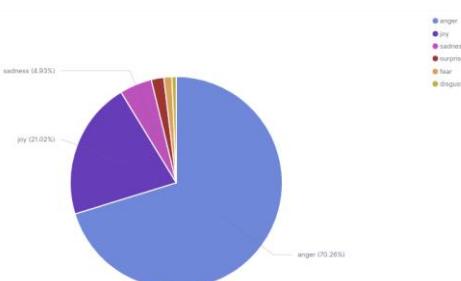
**Gambar 9. Word Matching**

#### B. Naive Bayes Emotion

Dengan memanfaatkan pustaka *Sentiment* pada R penggunaan Naive Bayes cenderung lebih mudah dan otomatis. *Bayes emotion* menghasilkan ekspresi pada setiap kalimat yang ada pada komentar seperti marah, santai, sedih, datar. hasil Naive Bayes Emotion terlihat pada **Gambar 10**, dan **Gambar 11**.

Anger	Disgust	Fear	Joy	Sadness	Surprise	Best Fit	Count
7341	3.092	2.068	1.025	1.728	2.787	anger	1.951
7341	3.092	2.068	1.025	1.728	7.341	anger	42
7341	3.092	2.068	1.025	1.728	11.895	surprise	1
7341	3.092	2.068	1.025	7.341	2.787	anger	20
7341	3.092	2.068	1.025	7.341	2.787	anger	1
7341	3.092	2.068	7.341	1.728	2.787	anger	121
7341	3.092	2.068	7.341	1.728	7.341	anger	1
7341	3.092	2.068	7.341	1.728	11.895	surprise	1
7341	3.092	2.068	7.341	7.341	2.787	anger	5
7341	3.092	2.068	13.656	1.728	2.787	joy	11
7341	3.092	2.068	13.656	7.341	2.787	joy	3
7341	3.092	2.068	13.656	1.728	2.787	joy	2
7341	3.092	2.068	19.972	1.728	2.787	joy	2
7341	3.092	2.068	38.818	1.728	2.787	joy	1

**Gambar 10. Tabel Naive Bayes Emotion**



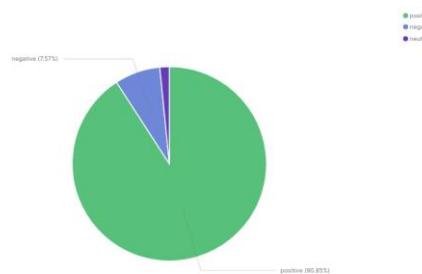
**Gambar 11. Diagram Naive Bayes Emotion**

#### C. Naive Bayes Polarity

Naive Bayes Polarity hanya menghasilkan klasifikasi berupa kategori kalimat apakah positif, negatif, atau netral. hasil dari klasifikasi *Polarity* terlihat pada **Gambar 12**, dan **Gambar 13**.

Negative	Positive	POS/NEG	Best Fit	Count
0.445	32.035	71.917	positive	1
0.445	34.808	78.141	positive	1
0.445	40.48	90.873	positive	1
0.445	42.559	95.541	positive	1
0.475	1.031	0.059	negative	871
0.475	8.782	0.927	negative	119
0.475	9.475	1	neutral	108
0.475	17.227	1.818	neutral	31
0.475	16.533	1.745	neutral	13
0.475	17.92	1.891	neutral	11
0.475	24.978	2.638	positive	6
0.475	25.671	2.709	positive	4
0.475	28.364	2.782	positive	4
0.475	34.115	3.6	positive	3
0.475	32.729	3.454	positive	2

**Gambar 12. Tabel Polarity**



**Gambar 13. Diagram Polarity**

#### V. KESIMPULAN

Penelitian perbandingan 2 metode untuk melakukan sentimen analisis menggunakan R. metode yang digunakan pada penelitian ini adalah metode *word matching* atau pencocokan sederhana dengan kamus yang ada, dan menggunakan algoritma *Naive Bayes*. Data yang digunakan berasal dari data komentar pada akun Instagram Tirto ID.

Hasil menunjukkan *Naive Bayes* pada pustaka sentimen R dapat menghasilkan dua jenis klasifikasi yaitu klasifikasi emosi seperti *ANGER*, *JOY*, *SAD*. dan klasifikasi *Polarity* berdasarkan negatif, positif dan netral. meskipun *Naive Bayes* bisa menghasilkan kecepatan dan akurasi yang tinggi namun metode *bayes* pada pustaka R terdapat kesalahan penterjemahan kalimat.

Untuk penelitian selanjutnya, akurasi dapat ditingkatkan lagi dengan melakukan data *preprocessing* yang lebih bersih dan pengecekan kosa kata, konteks kalimat, dan dialek daerah.

#### DAFTAR PUSTAKA

- [1] KBBI, “Sentimen”, 2019. [Online]. Available : <https://kbki.web.id/sentime>
- [2] O. U. Press, “Sentiment Analysis,” 2019. [Online]. Available: [https://www.lexico.com/en/definition/sentiment\\_analysis](https://www.lexico.com/en/definition/sentiment_analysis)
- [3] B. Pang and L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” in Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004, p. 271.
- [4] A. R. Naradhipa and A. Purwarianti, “Sentiment classification for indonesian message in social media,” in Cloud Computing and Social Networking (ICCCSN), 2012 International Conference on. IEEE, 2012, pp. 1–5.

- [5] P. Gamallo, M. Garcia, and S. Fernández-Lanza, "Tass: A naive-bayes strategy for sentiment analysis on spanish tweets," in Workshop on Sentiment Analysis at SEPLN (TASS2013), 2013, pp. 126–132.
- [6] C. Manning, P. Raghavan, dan H. Schutze, "Language models for information retrieval in introduction to information retrieval(pp. 237 - 252)
- [7] A. Clark, "Preprocessing very noisy text", Proc. of Workshop on Shallow Processing of Large Corpora, 2003, pp. 12-22.
- [8] E. Lunando dan A. Purwarianti, "Indonesian social media sentiment analysis with sarcasm detection" Advanced Computer Science and Information System (ICACSIS), 2013 International Conference on IEEE, 2013, pp. 195-198
- [9] T. N. Fatyasona dan Fitra A. Bachtiar, "Classification Method Comparison on Indonesian Social Media Sentiment Analysis" in International Conference on Sustainable Information Engineering and Technology (SIET), 2017 International Conference on IEEE, 2017.
- [10] Wahid, D. H., & Azhari, S. N. Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 10(2), 207-218, 2016
- [11] Lu, Bing, Hu, Minqing, and Cheng, Junsheng. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web Conference (WWW-2005), May 10-14, Chiba, Japan, 2005.