

# Analisis Sentimen Untuk Respon Masyarakat Terhadap Universitas (Studi Kasus : Universitas Mercu Buana Yogyakarta)

Anief Fauzan Rozi <sup>1</sup>, Agus Sidiq Purnomo <sup>2</sup>

<sup>1</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta

<sup>2</sup> Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta  
Yogyakarta, Indonesia

e-mail: <sup>1</sup> anief@mercubuana-yogya.ac.id, <sup>2</sup> sidiq@mercubuana-yogya.ac.id

Diajukan: 19 November 2020; Direvisi: 14 April 2021; Diterima: 04 Mei 2021

## ABSTRAK

*Dalam menjalankan tugasnya sebagai penghasil lulusan yang siap kerja dan digunakan oleh industri, tentunya Universitas Mercu Buana Yogyakarta (UMBY) juga harus menerapkannya secara menyeluruh, termasuk juga dalam upaya mengetahui respon atau mendapat masukan dari masyarakat terkait dengan eksistensi UMBY. Sehingga dalam penelitian ini dirumuskan bagaimana membuat peta analisis sentimen yang berisikan respon/feedback/komentar masyarakat dalam media sosial terkait dengan berbagai hal yang ada di UMBY. Pendekatan yang dilakukan dalam penyelesaian masalah dimulai dari pengumpulan data, preprocessing (cleansing, tokenizing, dan case folding), dan klasifikasi. Berdasarkan hasil penelitian yang telah dilakukan menunjukkan bahwa hasil sentimen analisis dengan menggunakan tagar #umby diperoleh data tweet sebanyak 179 data. Dengan hasil klasifikasi positif sebanyak 57 data (32%), negatif sebanyak 30 data (17%), dan netral sebanyak 92 data (51%).*

**Kata Kunci:** Analisis Sentimen, Twitter, Text Mining, UMBY

## ABSTRACT

*In carrying out its duties as a producer of graduates who are ready to work and used by industry, of course Mercu Buana Yogyakarta University (UMBY) must also apply it thoroughly, including in an effort to find out the response or get input from the public regarding the existence of UMBY. So that in this study formulated how to make a sentiment analysis map containing responses / feedback / comments from the public on social media related to various things that exist at UMBY. The approach taken in solving the problem starts from data collection, preprocessing (cleansing, tokenizing, and case folding), and classification. Based on the results of the research that has been done, it shows that the results of the sentiment analysis using the #umby hashtag obtained 179 tweet data. With the positive classification results of 57 data (32%), negative 30 data (17%), and neutral as many as 92 data (51%).*

**Keyword:** Sentiment Analysis, Twitter, Text Mining, UMBY

## 1. Pendahuluan

Siklus yang terjadi dalam Sistem Penjaminan Mutu Internal (SPMI) adalah Penetapan, Pelaksanaan, Evaluasi, Pengendalian, dan Peningkatan (PPEPP). Dalam menjalankan tugasnya sebagai pencetak lulusan yang siap kerja dan digunakan oleh industri, tentunya UMBY juga harus menerapkannya secara menyeluruh, termasuk juga dalam upaya mengetahui respon masyarakat terhadap berbagai hal. Salah satu hal yang dapat dilakukan adalah dengan mengetahui respon atau mendapat masukan dari masyarakat terkait dengan eksistensi UMBY. Mulai era 4.0 masyarakat cenderung menghabiskan waktunya dengan sosial media, twitter misalnya. Pada aplikasi tersebut masyarakat bisa saja memberikan tweet dengan *keyword* tertentu yang menyangkut dengan UMBY.

Teknik penarikan beberapa komentar tersebut dikenal dengan *Text Mining* yang difokuskan ke dalam *Sentiment Analysis*. Harapannya setelah ditemukannya peta analisis sentimen ini maka pihak UMBY dapat melakukan evaluasi dan tindak lanjut yang diperlukan.

Berdasarkan latar belakang tersebut, maka rumusan permasalahan dalam penelitian ini adalah bagaimana membuat peta analisis sentimen yang berisikan *respon/feedback/* komentar masyarakat dalam media sosial terkait dengan berbagai hal yang ada di UMBY, yang bertujuan untuk merancang peta analisis sentimen opini publik terhadap UMBY. Selanjutnya berdasarkan peta analisis tersebut akan dilakukan tindak lanjut yang dalam tahap ini masuk dalam langkah Evaluasi (PPEPP). Selanjutnya hasil dari penelitian ini diharapkan dapat digunakan oleh pihak UMBY sebagai masukan dan dapat melakukan tindak lanjut yang sesuai.

## 2. Tinjauan Pustaka

Beberapa penelitian yang terkait dalam penelitian ini antara lain seperti penelitian mengenai analisis sentimen *twitter* untuk teks berbahasa Indonesia dalam penelitian ini dilakukan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Data yang digunakan berasal dari jejaring sosial *Twitter* berdasarkan query dalam Bahasa Indonesia. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap objek tertentu yang disampaikan di *Twitter* dalam bahasa Indonesia, sehingga membantu usaha untuk melakukan riset pasar atas opini publik. Dalam penelitian ini digunakan Algoritma *Maximum Entropy* untuk POS *tagger* dan *Support Vector Machine* untuk membangun model klasifikasi atas data pelatihan. Hasil dari implementasi klasifikasi diperoleh akurasi 86,81 % pada pengujian *7 fold cross validation* untuk tipe kernel *Sigmoid*. Pelabelan kelas secara manual dengan POS *tagger* menghasilkan akurasi 81,67% [1].

Selanjutnya penelitian mengenai *twitter sentiment analysis* terhadap *brand reputation*. Dalam penelitian ini dilakukan perhitungan reputasi produk XL Axiata, dan dibandingkan dengan produk Telkomsel dan Indosat. Metode yang dilakukan dengan ekstrak data, analisis sentimen. Hasil dalam penelitian ini menunjukkan bahwa rata-rata perbandingan produk, XL Axiata memperoleh rata-rata nilai reputasi melebihi dibandingkan dengan Telkomsel dan Indosat yaitu sebesar 24.5%, sedangkan Telkomsel 13.2% dan Indosat 19.3% [2].

Selanjutnya penelitian mengenai Sentimen Twitter di SAMSAT. Penelitian ini bertujuan untuk membangun aplikasi analisis sentimen yang menerapkan pendekatan *Naïve Bayes Classifier* untuk mengklasifikasikan kata-kata dan difokuskan pada *tweet* dalam bahasa Indonesia. Data diperoleh melalui cara web scrapping dan sumber teks yang digunakan sebagai topik bahasan adalah Sistem Administrasi Manunggal Satu Atap (SAMSAT) Malang Kota. Proses klasifikasi dilakukan melalui serangkaian tahapan seperti preproses (*case folding, cleaning, tokenizing, dan stopword*) serta proses klasifikasi dengan algoritma *Naïve Bayes Classifier* itu sendiri untuk mendapatkan hasil klasifikasi dengan kategori positif, negatif atau netral. Hasil dari penelitian ini menunjukkan bahwa algoritma *Naïve Bayes Classifier* memberikan unjuk kerja yang baik dalam analisis sentimen. Dari hasil uji akurasi klasifikasi yang dilakukan oleh aplikasi menghasilkan nilai akurasi tertinggi pada setiap kategori positif, negatif, netral masing-masing sebesar 82%, 92%, 80% dengan jumlah data latih 200 *tweet* negatif, 200 *tweet* positif, dan 200 *tweet* netral [3].

Selanjutnya penelitian mengenai analisis sentimen terhadap pemerintahan Joko Widodo. Dalam penelitian ini analisis sentimen dilakukan dengan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Data opini diperoleh dari jejaring sosial *Twitter* berupa *tweet* berdasarkan *query* dalam Bahasa Indonesia. Data yang digunakan berjumlah 400 *tweet* terdiri dari 300 data latih dan 100 data uji. Data latih merupakan data yang telah diketahui sentimennya, 300 data latih terdiri dari 150 data kelas sentimen negatif dan 150 data kelas sentimen positif. Penelitian ini bertujuan untuk menentukan *tweet* merupakan *tweet* positif atau negatif yang disampaikan di *Twitter* dalam Bahasa Indonesia. Pengklasifikasian data *tweet* menggunakan algoritma *naïve bayes classifier*. Hasil klasifikasi pada data uji menunjukkan, algoritma *Naïve Bayes Classifier* memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentimen negatif dan 98% untuk sentimen positif [4].

Selanjutnya penelitian mengenai *sentiment analysis* terhadap *tweet* bernada sarkasme berbahasa Indonesia. pada penelitian ini dilakukan dibahas mengenai sentimen analisis bernada sarkasme pada *Tweet* berbahasa Indonesia dengan menggunakan fitur interjeksi dan unigram sebagai fitur utama mendeteksi kalimat sarkasme serta membandingkan 2 metode klasifikasi yaitu *Naïve Bayes* dan *Support Vector Machine* dengan kernel polinomial. Hasil dari penelitian ini menunjukkan penggunaan fitur interjeksi dan unigram sebagai pendeteksian sarkasme pada *tweet* berbahasa Indonesia mampu meningkatkan akurasi dengan rata-rata kenaikan akurasi lebih dari 8% untuk *classifier Naïve Bayes* dan lebih dari 13% untuk *classifier Support Vector Machine* dibandingkan hanya menggunakan fitur *unigram* saja. Hasil lainnya adalah akurasi terbaik adalah metode *Naïve Bayes* dengan akurasi terbaik yang diperoleh mencapai lebih dari 91 [5].

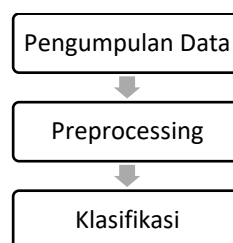
Selanjutnya penelitian mengenai sentimen analisis *twitter* terhadap suatu kebijakan pemerintah. Dalam penelitian ini dilakukan proses data mining (*transform, tokenize, stemming, dan classification*). Penelitian ini bertujuan untuk mengklasifikasikan sentimen positif, netral dan negatif dari *twitter* terhadap dua kata kunci yaitu “Jokowi” dan “Covid”. Dalam penelitian digunakan metode *Naive Bayes, Support Vector Machine, dan k-NN* untuk klasifikasi, dengan nilai terbaik yang dihasilkan yaitu *Support Vector Machine* dengan *accuracy* 84.58%, *precision* 82.14% dan *recall* 85.82% [6].

### 3. Metode Penelitian

*Sentimen analysis* adalah salah satu cabang penelitian *text mining*. Analisis sentimen berkaitan dengan bidang yang lebih luas seperti pengolahan bahasa alami, komputasi linguistik, dan *text mining* untuk tujuan menganalisis, sentiment, pendapat, sikap, evaluasi, penilaian dan emosi seseorang berkenaan pada topik, produk, layanan, organisasi, atau kegiatan tertentu [7] [8] [9].

Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif, negatif atau netral [8].

Secara garis besar tahap-tahap dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

#### 3.1 Pengumpulan Data

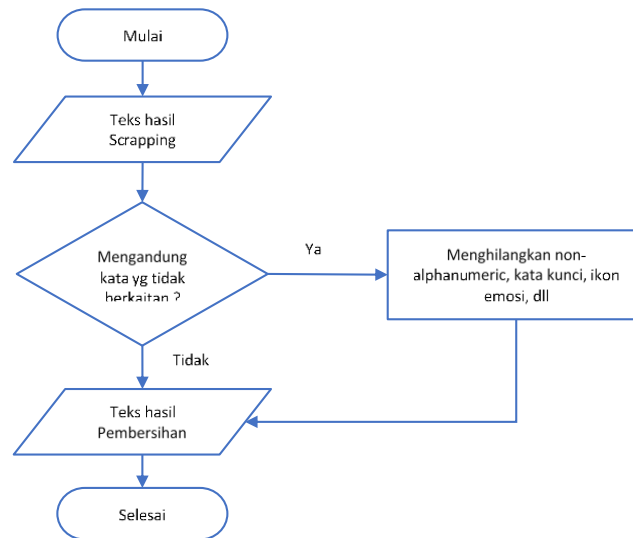
Data yang akan digunakan dalam penelitian adalah public timeline *tweet* yang merupakan hasil pencarian secara acak berdasarkan percakapan seseorang terhadap Universitas Mercu Buana Yogyakarta (UMBY) yang mengandung tagar *#umby*. Selanjutnya data didapatkan dengan *cara streaming* memanfaatkan API dari Twitter kemudian disimpan dalam bentuk *database*.

#### 3.2 Preprocessing

Pada tahap *preprocessing*, *tweet* yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri dari *cleansing, tokenizing, case folding*. Tujuan dari pemrosesan teks adalah supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan data.

##### 1. Cleansing

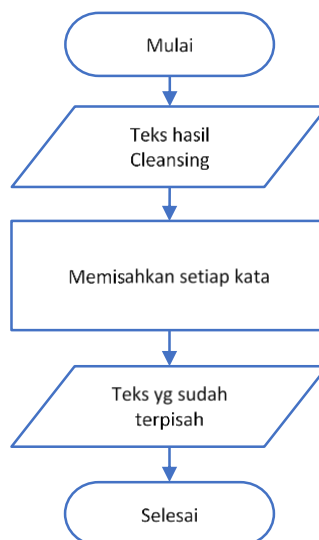
Teks hasil *scrapping* dari *twitter* setelah tahap akuisisi data selanjutnya akan dilakukan proses *cleansing* untuk menghilangkan karakter HTML, ikon emosi, *username, url, surat elektronik*. *Cleansing* bertujuan untuk meringankan beban dalam proses komputasi dan juga menghemat waktu. Proses *cleansing* dapat dilihat pada Gambar 2.



Gambar 2. Proses Cleansing

2. *Tokenizing*

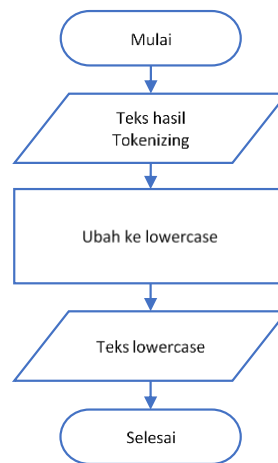
Pada tahap *tokenization*, teks hasil *cleansing* yang berarti sudah bersih karakter-karakter yang mengganggu akan dipisahkan untuk masing-masing kata pada setiap kalimat. Proses *tokenizing* dapat dilihat pada Gambar 3.



Gambar 3. Proses Tokenizing

3. *Case folding*

Teks yang telah melewati tahap *cleansing* dan *tokenization* selanjutnya akan dilakukan normalisasi, pada penelitian ini dilakukan normalisasi dengan cara merubah semua huruf menjadi huruf kecil. Tujuan dilakukan normalisasi tersebut adalah untuk memudahkan proses komputasi. Proses *case folding* dapat dilihat pada Gambar 4.



Gambar 4. Proses Case Folding

**3.3 Klasifikasi**

Pada hahap ini dilakukan proses pengklasifikasian untuk menentukan data mana yang termasuk dalam opini positif, negatif dan netral.

**4. Hasil Dan Pembahasan**

Pada saat proses akuisisi data teks yang diambil merupakan hasil scrapping sehingga memungkinkan banyak sekali karakter yang tidak digunakan ikut terambil. *Tweet* yang digunakan pada penelitian ini adalah *tweet* yang mengandung tagar #umby yang diambil secara acak berjumlah 179 *tweet*.

Teks yang didapatkan dari hasil akuisisi data merupakan data yang masih sangat mentah sehingga memerlukan beberapa tahapan lagi sebelum dapat diklasifikasikan pada proses *preprocessing* antara lain *cleansing*, *tokenizing* dan *case folding*.

Selanjutnya hasil dari *preprocessing* akan dilakukan klasifikasi dengan mencocokkan bank kata positif (3189 kata) dan negatif (7188 kata) untuk mendapatkan hasil akhir.

**4.1 Preprocessing**

**1. Cleansing**

*Cleansing* dilakukan untuk menghilangkan delimiter koma, titik, seluruh tanda baca, angka dalam *tweet* dan beberapa komponen yang ada di dalam *tweet* yaitu username, URL, karakter HTML, dan tagar karena tidak memiliki pengaruh apapun dalam proses analisis sentimen, maka komponen tersebut akan dihilangkan dengan tujuan untuk mengurangi *noise*. Adapun contoh *cleansing* data dari cuitan dapat dilihat pada Tabel 1.

Tabel 1. Contoh Data Cleansing

No	Text Tweets	Hasil Cleansing
1	Injap Semilunar Aorta: @swantemyid #PSLS #UMBY	Injap Semilunar Aorta
2	umby update hes much more alert today hes begun his 4 new daily meds hes a bit less yellow and ate his whole dinner portion yesterday #umby	umby update hes much more alert today hes begun his 4 new daily meds hes a bit less yellow and ate his whole dinner portion yesterday

**2. Tokenizing**

*Tokenizing* adalah proses yang dilakukan untuk memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau *temmed word*. Pada saat yang bersamaan, *tokenizing* juga membuang beberapa karakter tertentu yang dianggap sebagai tanda baca. Adapun contoh *tokenizing* dari cuitan dapat dilihat pada Tabel 2.

Tabel 2. Contoh Tokenisasi

No	Text Tweets	Hasil Tokenisasi
1	Injap semilunar aorta psls umby ruralhall httpstcomcahzberki	['Injap', 'semilunar', 'aorta', 'psls', 'umby', 'ruralhall', 'httpstcomcahzberki']
2	Yakali nder kalo umy umi kalo umby umbi gitu 🤔 😊	['Yakali', 'nder', 'kalo', 'umy', 'umi', 'kalo', 'umby', 'umbi', 'gitu 🤔 😊']

**3. Case Folding**

*Case Folding* bertujuan untuk mengubah seluruh ukuran objek pada kata menjadi ukuran yang sama. Karena tidak semua cuitan konsisten dalam menggunakan ukuran huruf. *Case folding* dilakukan dengan mengubah seluruh kata menjadi huruf kecil. Contoh dari penerapan *case folding* dapat dilihat pada Tabel 3.

Tabel 3. Contoh Case Folding

No	Text Tweets	Hasil Case Folding
1	['Injap', 'semilunar', 'aorta', 'psls', 'umby', 'ruralhall', 'httpstcomcahzberki']	['injap', 'semilunar', 'aorta', 'psls', 'umby', 'ruralhall', 'httpstcomcahzberki']
2	['Yakali', 'nder', 'kalo', 'umy', 'umi', 'kalo', 'umby', 'umbi', 'gitu 🤔 😊']	['yakali', 'nder', 'kalo', 'umy', 'umi', 'kalo', 'umby', 'umbi', 'gitu 🤔 😊']

**4.2 Klasifikasi**

Pada proses klasifikasi dalam penelitian ini belum menggunakan metode klasifikasi khusus, proses yang dilakukan hanya memeriksa term positif/negatif/netral pada *tweet* saja.

Data kemudian dibagi ke dalam tiga kelas, yaitu kelas positif, kelas negatif, dan kelas netral. pengambilan kelas berdasarkan kata yang sudah di dibandingkan dengan bank kata positif dan negatif yang telah di persiapkan, berikut ini contoh data yang sudah seperti pada Tabel 4.

Tabel 4. Contoh Klasifikasi

No	Text Tweets	Positive	Negative	Netral
1	umby is at the vet send good vibes his way please hes so sick rn	['good']	['sick']	['umby', 'is', 'at', 'the', 'vet', 'send', 'vibes', 'his', 'way', 'please', 'hes', 'so', 'rn']
2	umby probably the new champion delay that year but tbh idk	['champion']	['delay']	['umby', 'probably', 'the', 'new', 'that', 'year', 'but', 'tbh', 'idk']

No	Text Tweets	Positive	Negative	Netral
3	disini ada yg anak umby ngga menurut kalian mending kampus 1 atau kampus 2 yaa untuk psikologi dari segi tempatnya dll juga cm	[]	[]	['disini', 'ada', 'yg', 'anak', 'umby', 'ngga', 'menurut', 'kalian', 'mending', 'kampus 1', 'atau', 'kampus 2', 'yaa', 'untuk', 'psikologi', 'dari segi', 'tempatnya', 'dll', 'juga', 'cm']

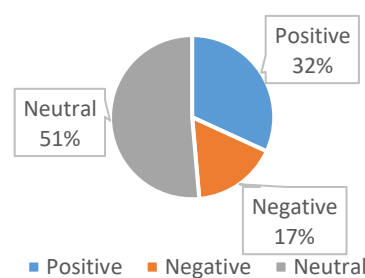
Berdasarkan hasil klasifikasi data yang diperoleh dari 179 tweet, setelah diklasifikasikan ke dalam kategori positif, negatif, dan netral seperti pada Tabel 4. Proses selanjutnya dilakukan perhitungan jumlah tweet seperti pada Tabel 5.

Tabel 5. Jumlah Tweet

Positive	Negative	Neutral
57	30	92

Berikut ini hasil *output* analisis dalam bentuk pie diagram yang menyajikan persentase diagram/persentase positif, negatif, dan netral dapat dilihat pada Gambar 5.

Hasil Analisis (Berdasarkan Jumlah Tweet)



Gambar 5. Hasil Analisis

### 5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan menunjukkan bahwa hasil sentimen analisis dengan menggunakan tagar #umby diperoleh data tweet sebanyak 179 data. Dengan hasil klasifikasi positif sebanyak 57 data (32%), negatif sebanyak 30 data (17%), dan netral sebanyak 92 data (51%).

Saran untuk penelitian selanjutnya antara lain sebagai berikut :

1. Hasil pemrosesan data dapat diintegrasikan menggunakan *query* SQL, sehingga dapat divisualisasikan menggunakan web.
2. Klasifikasi dapat dilanjutkan menggunakan beberapa metode tambahan.

---

### Daftar Pustaka

- [1] N. D. Putranti dan E. Winarko, “Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine,” *IJCCS, Vol.8, No. 1, ISSN : 1978-1520*, pp. 91-100, 2014.
- [2] N. A. Vidya, “Twitter Sentimen Analysis Terhadap Brand Reputation : Studi Kasus PT XL Axiata Tbk,” Universitas Indonesia, Jakarta, 2015.
- [3] I. F. Rozi, E. N. Hamdana dan M. B. I. Alfahmi, “Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus SAMSAT Kota Malang),” *Jurnal Informatika Polinema, Vol. 4, Ed. 2, ISSN : 2407-070X*, pp. 149-154, 2018.
- [4] Y. S. Mahardhika dan E. Zuliarso, “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier,” dalam *SINTAK, ISBN: 978-602-8557-20-7*, Semarang, 2018.
- [5] L. Septiani dan Y. Sibaroni, “Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia,” *Jurnal Linguistik Komputasional, Vol. 2, No. 2, e-ISSN : 2621-9336*, pp. 62-67, 2019.
- [6] S. Hikmawan, A. Pardamean dan S. N. Khasanah, “Sentimen Analisis Publik Terhadap Joko Widodo Terhadap Wabah Covid-19 Menggunakan Metode Machine Learning,” *Jurnal Kajian Ilmiah (JKI), Vol. 20 No. 2 (Mei 2020), e-ISSN: 2597-792X, ISSN: 1410-9794*, pp. 167-176, 2020.
- [7] B. Liu, *Sentiment Analysis and Opinion Mining*, California: Morgan & Claypool, 2012.
- [8] V. A. Kharde dan S. S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques,” *International Journal of Computer Applications*, vol. 139, no. 11, pp. 5 - 15, 2016.
- [9] A. Rasool, R. Tao, K. Marjan dan T. Naveed, “Twitter Sentiment Analysis: A Case Study for Apparel Brands,” *IOP Conf. Series: Journal of Physics*, vol. 1176, pp. 1 - 6, 2019.
- [10] K. Dave, S. Lawrence dan D. M. Pennock, “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews,” Budapest, Hungary, 2003.