



Sentiment Analysis Against Beauty Shaming Comments on Twitter Social Media Using SentiStrength Algorithm

Analisis Sentimen Terhadap Komentar Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength

Septika Sari^{1*}, Pradita Eko Prasetyo Utomo, S.pd., M.Cs.², Ulfa Khaira, S.Komp., M.Kom.³,
Tri Suratno, S.Kom., M.Kom.⁴

Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

E-Mail: ¹tsari722@gmail.com, ²Pradita.eko@unj.ac.id, ³ulfa.ilkom@gmail.com, ⁴trie1123@gmail.com

Makalah: Diterima 23 Februari 2021; Diperbaiki 01 Maret 2021; Disetujui 31 Maret 2021
Corresponding Author: Septika Sari

Abstrak

Perkembangan teknologi saat ini khususnya di bidang teknologi informasi telah mengalami kemajuan dan perkembangan yang pesat, baik di bidang jaringan komunikasi, *Hardware* maupun *Software* serta infrastruktur lainnya. Media sosial merupakan salah satu bentuk perkembangan dari jaringan internet yang dapat diakses secara *online*. Penggunaan media sosial didalam kehidupan sehari-hari telah membawa pengaruh besar dalam perubahan aspek kehidupan. Salah satu media sosial yang sedang populer dan digunakan oleh banyak masyarakat adalah *Twitter*. Penggunaan media sosial *Twitter* digunakan untuk berbagai kepentingan seperti ajang promosi atau hanya sekedar memberikan *tweet* berupa opini atau fakta. Namun tidak jarang, *tweet* yang diberikan pengguna *Twitter* mengandung unsur negatif seperti *Beauty Shaming*. Analisis sentimen dilakukan untuk mengidentifikasi atau memahami unsur *Beauty Shaming* dari suatu data *tweet*. Metode penelitian yang digunakan adalah Algoritma *SentiStrength* untuk analisis sentimen. Data *tweet* yang didapatkan berjumlah 300 *tweets* dari proses *crawling* data. Hasil analisis sentimen tersebut, didapatkanlah 93 *tweet Beauty Shaming* (33,7%), 117 *tweet* netral (43,2%), dan 63 *tweet non-Beauty Shaming* (23,1%) dari 273 *tweet* pengguna.

Keyword: Analisis Sentimen, *Beauty Shaming*, Klasifikasi, *Sentistrength*, *Twitter*.

Abstract

Current technological developments especially in the field of information technology, have progressed and developed rapidly, both in the field of communication networks, hardware and software and other infrastructure. Social media is a form of development of internet networks that can be accessed online. The use of social media in everyday life has had a major influence in changing aspects of life. One of the social media that is currently popular and used by many people is *Twitter*. The use of *Twitter* social media is used for various purposes such as promotional events or just to provide tweets in the form of opinions or facts. But not infrequently, tweets provided by *Twitter* users contain negative elements such as *Beauty Shaming*. Sentiment analysis is carried out to identify or understand the *Beauty Shaming* element of a tweet data. The research method used is the *SentiStrength* Algorithm for sentiment analysis. The tweet data obtained were 300 tweets from the data crawling process. The results of the sentiment analysis revealed 93 *Beauty Shaming* tweets (33.7%), 117 neutral tweets (43.2%), and 63 non-*Beauty Shaming* tweets (23.1%) from 273 user tweets.

Keyword: *Beauty Shaming*, Classification, Sentiment Analysis, *Sentistrength*, *Twitter*

1. Pendahuluan

Perkembangan teknologi saat ini khususnya di bidang teknologi informasi telah mengalami kemajuan dan perkembangan yang pesat, baik di bidang jaringan komunikasi, *Hardware* maupun *Software* serta infrastruktur lainnya yang dapat mendukung sebuah sistem informasi tersebut mengalami kemajuan dan perkembangan juga. Perkembangan jaringan komunikasi misalnya didalam penggunaan media internet sebagai sarana komunikasi pun mengalami kemajuan misalnya berkembangnya telepon seluler menjadi telepon cerdas atau

smartphone. Terciptanya *smartphone*, layanan untuk pengguna yang disediakan dalam berkomunikasi pun semakin beraneka ragam, mulai dari fitur *chatting*, *browsing* serta layanan media sosial.

Media sosial salah satu perkembangan jaringan internet yang diakses secara online dimana para penggunanya dapat saling berkomunikasi satu sama lain dan memungkinkan mendapatkan berbagai informasi dari berbagai bidang. Penggunaan media sosial terhadap kehidupan serta kegiatan masyarakat sehari-hari telah membawa pengaruh dalam perubahan kebudayaan, gaya hidup, bahasa serta pergaulan masyarakat[1]. Salah satu contoh bentuk media sosial yang sedang populer saat ini adalah *Twitter*. *Twitter* adalah salah satu media sosial hasil pemikiran Dorsey ini yang sangat diminati saat ini termasuk negara Indonesia. Minatnya masyarakat Indonesia terhadap *Twitter* dapat menjadi salah satu alasan mengapa banyak pula kalangan yang menggunakan media sosial tersebut untuk berbagai kepentingan seperti berbagi informasi pemberitaan, ajang promosi dan bisnis atau hanya sekedar tempat memberikan *tweet*, komentar, fakta ataupun opini[2]. Namun tidak jarang bahwa pengguna *Twitter* memberikan *tweet*, komentar, fakta ataupun opini yang tidak mengenakan dan cenderung menyinggung pengguna *Twitter* lain dengan kata-kata atau kalimat yang mengandung ke arah yang negatif. Salah satu bentuk *tweet* atau komentar negatif yang sering digunakan pengguna *Twitter* adalah *Beauty Shaming*.

Secara umum, *Beauty Shaming* bisa diartikan sebagai kegiatan atau perilaku menjelek-jelekan atau mengomentari penampilan fisik seseorang seperti mengomentari fisik wajah, dan bentuk badan. Perilaku ini bisa terjadi di kalangan mana pun, pria maupun wanita. Alasan pengguna *Twitter* melakukan *Beauty Shaming* pun beragam seperti sebagai bahan canda tawa, mencairkan suasana, sekadar iseng atau ikut-ikutan pengguna *Twitter* lain, hingga memang ingin menjelek-jelekan seseorang. *Beauty Shaming* bukanlah perilaku yang bisa dianggap remeh. Dampak dari *Beauty Shaming* yaitu menurunkan rasa percaya diri, menimbulkan gangguan depresi, dan yang lebih fatalnya dapat meningkatkan risiko bunuh diri pada korban *Beauty Shaming*. Perilaku tersebut jika terus menerus berlanjut akan menimbulkan hubungan signifikan antara yang melakukan *Beauty Shaming* dan korban *Beauty Shaming* yang mana semakin reaktif perilaku *Beauty Shaming* maka semakin reaktif pula perilaku korban *Beauty Shaming*[3].

Perilaku pengguna *Twitter* yang berbentuk opini maupun fakta terhadap *Beauty Shaming* jika diolah dengan baik dan benar dapat menjadi sumber data yang menghasilkan informasi yang digunakan untuk meningkatkan persepsi seseorang terhadap hal tersebut. Bentuk analisis yang dapat dilakukan untuk menganalisis perilaku tersebut adalah analisis sentimen. Analisis sentimen merupakan suatu proses yang dilakukan untuk memahami, mengekstrak, dan mengolah data berupa tekstual secara otomatis yang digunakan untuk mendapatkan informasi[4]. Salah satunya dengan menganalisis sentimen pengguna *Twitter* dengan menggunakan hasil *tweets* dari pengguna *Twitter* yang melakukan *Beauty Shaming*. Penelitian terkait analisis sentimen dari media sosial *Twitter* telah banyak dilakukan sebelumnya dengan metode analisis yang berbeda-beda. Pada penelitian yang dilakukan oleh Sianipar R dan Setiawan pada tahun 2015 yang berjudul "Pendeteksian Kekuatan Sentimen pada Teks *Tweet* Berbahasa Indonesia Menggunakan *SentiStrength*".

Metode *SentiStrength* dievaluasi pada beberapa dataset yang terdiri dari teks bahasa Indonesia. Evaluasi metode *SentiStrength* yang dilakukan peneliti tersebut adalah akurasi, uji korelasi, dan uji hipotesis. *SentiStrength* pada penelitian tersebut menggunakan pendekatan berbasis kamus atau leksikon yang berisi *terms* serta bobot kekuatan sentimennya. Dikarenakan kamus atau leksikon default *SentiStrength* dibuat dalam bahasa Inggris, maka kamus leksikon telah diterjemah dalam bahasa Indonesia untuk kebutuhan penelitian tersebut[5]. Penelitian ini mengadaptasi *classifier SentiStrength* yang sama seperti penelitian tersebut dan menggunakan leksikon yang telah diterjemah dalam bahasa Indonesia.

2. Metodologi Penelitian

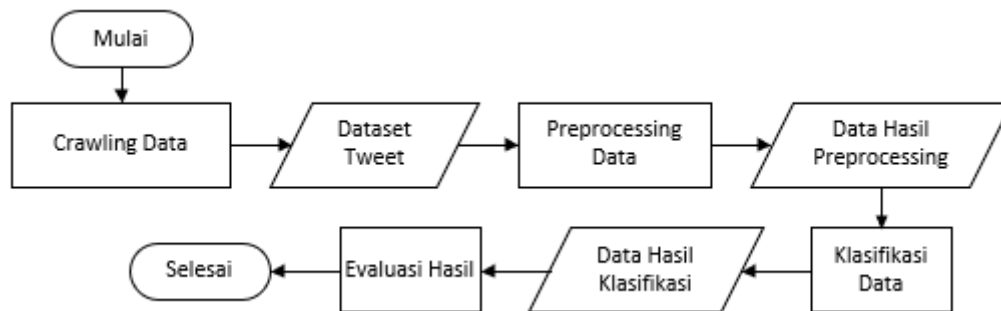
Suatu penelitian dimulai dengan perencanaan yang seksama dengan mengikuti petunjuk yang disusun secara logis dan sistematis, sehingga hasilnya dapat menjalankan kondisi yang dijelaskan. Metode penelitian yang dilakukan menggunakan Algoritma *SentiStrength*. Algoritma *SentiStrength* merupakan Algoritma untuk *opinion mining* yang menggunakan pendekatan berbasis kamus atau leksikon yang bekerja dengan mendeteksi setiap kata atau frasa dari suatu teks dengan mengecek *terms* yang mengandung sentimen dan kemudian *output* yang dihasilkan adalah bobot kata atau frasa yang berhasil terdeteksi, masing-masing kelas adalah kelas positif, negatif dan netral.

2.1 Prosedur Penelitian

Prosedur penelitian menjelaskan tahapan-tahapan yang akan dilakukan dalam penelitian untuk mendapatkan hasil akhir dari analisis sentimen terkait *Beauty Shaming* dari hasil *tweet* dari pengguna *Twitter*. Percobaan penelitian dimulai dari pengambilan data *tweet* dari *Twitter* melalui proses yang dikenal dengan istilah *crawling*, terdiri dari beberapa penggunaan *keyword*. Kemudian data hasil *crawling* tersebut dilakukan teks

preprocessing atau pemrosesan teks, yang terdiri dari beberapa tahapan yaitu *cleaning*, *tokenization*, *stopword*, dan *stemming*.

Setelah melakukan pemrosesan teks atau teks *processing*, kemudian dari hasil pemrosesan tersebut dilakukan klasifikasi data dengan metode algoritma *SentiStrength* untuk mengidentifikasi kata yang mengandung *Beauty Shaming*, *non-Beauty Shaming*, ataupun netral. Untuk mengakhiri tahapan percobaan penelitian, dilakukan evaluasi hasil klasifikasi data *tweet* tersebut. Prosedur percobaan penelitian berupa *flowchart* dapat dilihat pada Gambar 1.



Gambar 1. Prosedur Penelitian

2.2 Crawling Data

Dataset yang digunakan untuk penelitian ini merupakan data *tweet* dari *Twitter* yang telah melalui proses *crawling* data menggunakan bantuan *Application Programming Interface (API) Twitter*. *Crawling* data merupakan suatu proses untuk mengambil atau mengunduh data dari *server Twitter* berupa data pengguna atau data *tweet* sesuai *keyword* yang digunakan untuk penelitian ini[6]. Proses dilakukannya *crawling* data adalah membuat program dengan menggunakan kata kunci atau *keyword* untuk mendapat data *tweet* sesuai dengan yang diinginkan. *Keyword* yang digunakan dalam pengambilan data *tweet* merupakan frasa atau kalimat yang cenderung mengandung maksud buruk atau unsur menjelek-jelekan penampilan fisik. Salah satu contohnya seperti “muka lu jelek banget”. Jumlah *tweet* yang didapatkan dari kata kunci atau *keyword* tersebut adalah sebanyak 300 *tweet* data awal.

Adapun *tools* yang digunakan dalam proses *crawling* data diantaranya: *plugin python-twitter*, *nlTK*, *credentials*, *pandas*, *numpy*, *sastrawi*, *unicode*, *textblob*, *matplotlib* dan *WordCloud*. Dibuat melalui IDE *Jupyter* dalam lingkungan bahasa pemrograman *Python*.

2.3 Preprocessing Data

Dataset yang digunakan berupa data *tweet* yang didapatkan dari proses *crawling* data masih berbentuk data yang belum terstruktur masih terdapat banyak simbol dan kata yang tidak perlu digunakan, karena itu perlu dilakukan tahapan *preprocessing* data untuk mendapatkan data bersih agar proses pengklasifikasian data lebih baik dan akurat. *Preprocessing* data adalah proses tahapan awal dari data *mining* dimana data yang telah didapatkan dilanjutkan dengan pengolahan data dengan tahapan-tahapan yang terdiri dari *proses cleaning*, *tokenization*, *stopword*, dan *stemming*[7].

2.3.1 Proses Cleaning

Cleaning data merupakan proses pembersihan dan pembuangan data dengan menghilangkan noise dan data yang tidak relevan dimana memiliki unsur yang tidak sempurna seperti missing value, mengandung emoticon, salah dalam pengetikan ataupun yang tidak valid. Proses ini akan mempengaruhi performa dari *text mining* yang dilakukan karena data yang digunakan akan berkurang jumlah serta kompleksitas dari data tersebut[8].

2.3.2 Proses Tokenization

Proses Tokenisasi merupakan proses tahapan pemisahan suatu rangkaian karakter berdasarkan tiap kata yang menyusunnya atau karakter spasi, dan kemungkinan di waktu yang bersamaan dilakukan penghapusan karakter kata. Dalam proses tokenisasi, dibutuhkan suatu *term* atau kata, sebagai contoh sebuah *token* merupakan suatu urutan karakter dari dokumen tertentu yang dikelompokkan sebagai satuan unsur makna yang berguna untuk membantu pemrosesan tokenisasi[9]. Sebelum dilakukannya proses tokenisasi, data *tweet* yang didapatkan tersebut harus telah melalui proses *cleaning*.

2.3.3 Proses Stopword

Proses *stopword* merupakan tahapan pemrosesan kata-kata umum yang sering namun tidak memberikan informasi yang penting dan biasanya diacuhkan atau dibuang. Contoh *stopword* dalam bahasa Indonesia antara

lain “yang”, “ke”, “di” dan sebagainya. Selanjutnya melalui proses *stopword removal* yaitu proses *filtering*, dimana kata-kata penting tersebut dipilih dari hasil *token* yaitu kata-kata apa yang digunakan yang berhubungan dengan data dokumen tersebut[10].

2.3.4 Proses Stemming

Proses *stemming* merupakan proses dimana dilakukannya pemetaan dan penguraian berbagai bentuk (*variants*) menjadi bentuk *stem* (kata dasar) dari kata hasil pemrosesan *stopword removal* atau *filtering*. Proses *stemming* digunakan untuk meningkatkan kualitas informasi dari data yang didapatkan. Terdapat dua aturan untuk melakukan proses *stemming* tersebut yaitu pendekatan kamus dan aturan. Sebagai contoh kata “mengerjakan”, “dikerjakan” dan “pekerja” yang semula mengandung arti yang berbeda dapat dilakukan stem menjadi sebuah kata “kerja” yang memiliki arti yang sama sehingga kata tersebut berhubungan[11].

2.4 Klasifikasi Data

Pada penelitian ini, klasifikasi data berupa data *tweet* menggunakan algoritma *SentiStrength*. *Sentistrength* merupakan algoritma klasifikasi yang menggunakan pendekatan leksikon dan penggunaan aturan-aturan informasi linguistik tambahan yang tidak leksikal untuk mengidentifikasi kekuatan sentimen di sebuah teks[12]. *SentiStrength* menggunakan sistem *two-scale* yakni positif dan negatif, dikarenakan menurut penelitian dibidang psikologi bahwa manusia dapat merasakan emosi positif dan negatif secara bersamaan hingga batas yang tertentu yang dibatasi secara mandiri[13]. *SentiStrength* akan menghasilkan nilai positif dan negatif, dimana nilai jangkauan tersebut dimulai dari angka 1 sampai 5. Nilai 1 menunjukkan pada kalimat yang tidak memiliki sentimen positif maupun negatif, sedangkan nilai 5 menunjukkan pada kalimat yang memiliki sentimen yang sangat positif atau sangat negatif.

Nilai akhir pada sebuah kalimat tersebut ditentukan oleh skor positif tertinggi dan skor negatif tertinggi dari kata-kata yang menyusunnya. Contohnya seperti : “Saya benar-benar menyayangimu tapi kamu tidak menyayangi saya”. Maka hasil *parsing* kalimat tersebut sebagai berikut : Saya benar-benar menyayangimu(4) tapi kamu tidak sayang(-3) saya <skor:4,-3>. Angka dalam (...) memunjukkan berapa angka kekuatan sentimen pada setiap kata penyusun, sedangkan angka di dalam tanda “<skor; ... , ...>” yang mengindikasikan skor akhir dari kalimat tersebut. Keputusan akhir hasil sentimen didasarkan pada aturan :

If positive value > negative value then positive sentiment.
If positive value < negative value then negative sentiment.
If positif value = negative value then neutral sentiment.

SentiStrength dapat dimodifikasikan untuk bahasa lainnya , seperti contoh Bahasa Indonesia dimana telah diterjemah sebelumnya dari isi leksikon *default*-nya, menyesuaikan aturannya untuk mengatasi fitur pada struktur jenis bahasa tertentu. Aturan yang ada di *SentiStrength* dapat dimodifikasikan bahkan dihilangkan sesuai kebutuhan dari bahasa tersebut. Pengaplikasian untuk bahasa selain bahasa Inggris cenderung menghasilkan niai akurasi yang lebih rendah dibanding bahasa inggris dikarenakan analisis dan prosesnya lebih sering fokus pada bahasa Inggris.

2.5 Uji Akurasi Klasifikasi Data

Pengujian klasifikasi sentimen pada data *tweet* tersebut dilakukan dengan membandingkan antara data prediksi serta data aktual. Data prediksi tersebut berupa hasil pengklasifikasian yang dihasilkan dari algoritma yang digunakan yaitu *SentiStrength*, sedangkan data aktual tersebut berupa hasil pengklasifikasian yang dihasilkan dari pelabelan *tweet* yang dilakukan secara manual ke dalam kelas positif, negatif, dan netral. Penggunaan algoritma *SentiStrength* tidak membutuhkan data latih untuk melakukan klasifikasi sentimen, karena *SentiStrength* merupakan algoritma yang bersifat *unsupervised* (klasifikasi tak terbimbing) sehingga pengujian tersebut dilakukan dengan membagi data uji pada data *tweet*.

Pada penelitian ini dilakukan pengujian data menggunakan uji akurasi secara intrinsik. Intrinsik adalah pengukuran kualitas suatu data prediksi yang nantinya akan dibandingkan dengan data aktual yang dilabelkan secara manual dengan menghitung akurasi. Uji akurasi tersebut menunjukkan perbandingan data yang terklasifikasi benar dan dibandingkan dengan jumlah semua data. Untuk menghitung tingkat akurasi dari klasifikasi tersebut digunakanlah metode *confusion matrix*, atau biasanya dikenal dengan istilah *matching matrix*. Tabel matriks terdiri dari data prediksi dan aktual dengan model *confusion matrix* dengan bentuk 3x3 [14]. Tabel tersebut dapat dilihat pada Tabel 1.

Tabel 1. Tabel *Confusion Matrix*

		Data Prediksi		
		Kelas A	Kelas B	Kelas C
Data Aktual	Kelas A	AA	AB	AC
	Kelas B	BA	BB	BC
	Kelas C	CA	CB	CC

Perhitungan uji akurasi *confusion matrix* ditunjukkan pada persamaan 1.

$$Akurasi = \frac{AA + BB + CC}{AA + AB + AC + BA + BB + BC + CA + CB + CC} \quad (1)$$

3. Hasil dan Analisis

3.1 Preprocessing Data

Preprocessing merupakan tahapan awal untuk melakukan *text mining* berupa mengubah data sesuai format yang dibutuhkan. Didalam proses ini dilakukan proses penggalian, pengolahan dan pengaturan informasi dimana akan digunakan untuk menganalisis bagaimana hubungan tekstual antara data yang terstruktur dan data yang tidak terstruktur. Tahapan *preprocessing* data terdiri dari proses *cleaning*, *tokenize*, *stopword*, dan *stemming*.

3.1.1 Proses Cleaning

Pada proses ini dilakukan suatu proses memperbaiki, mendeteksi serta pembersihan suatu record data *tweet* yang tidak lengkap, tidak terstruktur dan tidak relevan dengan keyword yang selanjutnya data *tweet* tersebut dimodifikasi dan mempermudah proses *preprocessing* data. Pembersihan data *tweet* meliputi penghapusan URL, *username*, *mention* pengguna, membuang duplikasi kata, memperbaiki kesalahan pada data seperti kesalahan dalam penulisan, serta simbol-simbol seperti tanda baca dan *emoticon*. Hasil pemrosesan *cleaning* data dapat dilihat pada Tabel 2.

Tabel 2. Proses *Cleaning*

Sebelum	Sesudah
@pnstol humor lu jelek banget.', 'mirip sama gue 😊 😊 😊 🗣️ 🗣️ 🗣️	humor lu jelek banget mirip sama gue

3.1.2 Proses Tokenization

Tokenisasi adalah proses pemisahan teks pada data menjadi potongan-potongan yang disebut *token* untuk kemudian dianalisa di proses selanjutnya dengan tujuan untuk memisahkan kalimat atau kata-kata pada data *tweet*. Hasil proses tokenisasi data *tweet* dapat dilihat pada Tabel 3.

Tabel 3. Proses *Tokenization*

Sebelum	Sesudah
humor lu jelek banget mirip sama gue	'humor', 'lu', 'jelek', 'banget', 'mirip', 'sama', 'gue'

3.1.3 Proses Stopword

Setelah melakukan tokenisasi pada data *tweet*, data hasil proses tokenisasi tersebut kembali dilakukan pemrosesan *stopword* dan *stopword removal*. *Stopword* merupakan proses pembuangan kata-kata umum yang sering muncul atau digunakan namun tidak memberikan informasi dan makna yang penting. Contohnya seperti di, yang, ke, jadi, sama, dan sebagainya. Dari hasil pemrosesan *stopword* didapatkanlah hasil data *tweet* seperti pada Tabel 4.

Tabel 4. Proses *Stopword*

Sebelum	Sesudah
'humor', 'lu', 'jelek', 'banget', 'mirip', 'sama', 'gue'	'humor', 'lu', 'jelek', 'banget', 'mirip', 'gue'

3.1.4 Proses Stemming

Proses *stemming* bertujuan untuk mentransformasikan kata dari data *tweet* menjadi kata dasarnya dengan menghilangkan atau membuang semua imbuhan awalan dan akhir dari kata pada data *tweet* tersebut. Dalam proses ini, semua kata pada data *tweet* tersebut menjadi huruf *lowercase* atau huruf kecil. Hasil pemrosesan *stemming* data dapat dilihat pada Tabel 5.

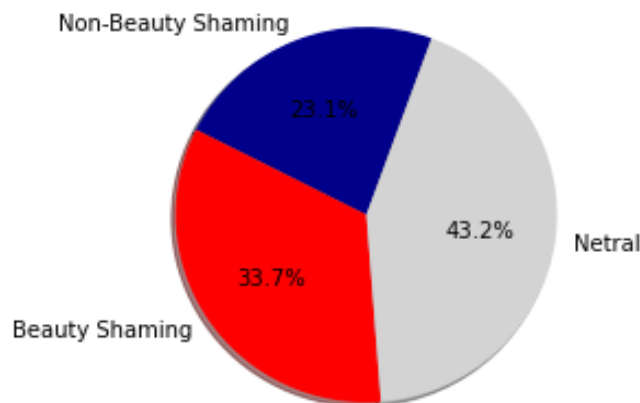
Tabel 5. Proses *Stemming*

Sebelum	Sesudah
‘humor, ‘lu’, ‘jelek’, ‘banget’, ‘mirip, ‘gue’	Humor lu jelek banget mirip gue

3.2 Klasifikasi Data

Pada proses klasifikasi data dilakukanlah klasifikasi menggunakan algoritma *SentiStrength* dan mengadaptasi *classifier SentiStrength* menggunakan leksikon yang telah diterjemah dalam bahasa Indonesia dan memanfaatkan penggunaan leksikon dengan sistem *two-scale* yakni positif dan negatif dan hasil selanjutnya akan menghasilkan nilai positif dan negatif. Proses klasifikasi data menggunakan kamus sentimen yang diperlukan untuk dijadikan sebagai patokan dalam menentukan sentimen yang terdapat didalam data *tweet* tersebut. Total data *tweet* asli hasil pengambilan data melalui proses *crawling* data berjumlah 300 *tweets*, dalam pengujian klasifikasi ini digunakan 273 *tweets* hasil *preprocessing*.

Pada uji klasifikasi, teks kata didalam data hasil *preprocessing* sebelumnya akan dibandingkan terhadap isi leksikon apakah isi data sama dengan *opinion words* di leksikon. Jika sama, akan ditampilkan bobot dari kata yang sama tersebut. Nilai akhir pada data *tweet* ditentukan dari nilai positif dan negatif tertinggi dari leksikon tersebut. Data *tweet* yang memiliki jumlah nilai positif lebih tinggi dari nilai negatif maka termasuk kedalam sentimen positif, begitupun sebaliknya dan jika data *tweet* tersebut memiliki jumlah nilai positif dan negatif yang sama, maka termasuk kedalam sentimen netral.



Gambar 2. Plot Hasil Analisis Sentimen

Gambar 2 menunjukkan hasil analisis sentimen dari 273 data *tweet* dari pengguna *Twitter*. Dengan persentase *Beauty Shaming* 33,7%, *Non-Behauty Shaming* 23,1% dan Netral 43,2%. Diperoleh dari uji klasifikasi tersebut adalah Netral yang paling tinggi atau banyak. Meskipun, kata kunci atau *keyword* yang digunakan mengandung frasa atau kalimat yang mengandung unsur *Beauty Shaming* atau jelek dan sebagainya, akan tetapi terdapat kemungkinan bahwa hasil yang diperoleh dapat bersifat netral.

3. Berdasarkan hasil pengujian tingkat akurasi diperoleh hasil dari tingkat akurasi klasifikasi data sebanyak 60%, karena terdapat perbedaan penilaian antara menggunakan algoritma dan manusia dalam melakukan analisis sentimen dari data *tweet*, hal ini dikarenakan manusia memiliki emosi jiwa dan rasa yang berbeda dalam memberikan seberapa penilaian dari sentimen data *tweet* tersebut.

Referensi

- [1] R. Nasrullah, *Media Sosial: Perspektif Komunikasi, Budaya, dan Sosioteknologivol*, Bandung : Simbiosis Rekatama Media, 2015.
- [2] A. Samana, *Ekspresi Diri Net Generation Melalui Twitter*, Universitas Airlangga : Pasca Sarjana, 2013.
- [3] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 10, no. 2, p. 207, Jul. 2016.
- [4] B. Pang, L. Lee, "Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval", Vol. 2, No 1-2 pp. 1-135, 2008.
- [5] R. Sianipar and E. B. Setiawan, "Pendeteksian Kekuatan Sentimen Pada Teks Tweet Berbahasa Indonesia Menggunakan Sentistrength," *e-Proceeding Eng.* Vol.2, No.3, Desember 2015.
- [6] Sembodo, J. E., Setiawan, E. B., & Baizal, Z. A, "Data Crawling Otomatis pada Twitter. In Indonesian Symposium on Computing (Indo-SC)", pp.11-16, 2016.
- [7] V. Srividhya, & R. Anitha, "Evaluating preprocessing techniques in text categorization. International journal of computer science and application", 47(11), 49-51, 2010.
- [8] A. Yuli, & R. Mardhiyah, "Aplikasi Data Minig untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa, Jurnal Informatika, vol. 8, no.1, Januari 2014.
- [9] Salton, G., *Automatic Text Processing, TheTransformation, Analysis, and Retrieval of information by computer*, Inc. USA : Addison – Wesly Publishing Company, 1989.
- [10] S.A. Muhammad, Imam M.I.S. and Sri M., "Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal dan Stemming Berbasis Android", *Prosiding Konferensi Ilmiah Mahasiswa UNISSULA (KIMU) 2*, Oktober 2019.
- [11] S.S.H. Manase, "Studi Perbandingan Algoritma-Algoritma Stemming untuk Dokumen Teks Bahasa Indonesia", *Jurnal INFOKAR*, Vol. 1, No.1, Juli 2017.
- [12] Thelwall, M., Buckley, K., Paltoglou et.al., "Sentiment Strength Detection in Short Informal Text", *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558, 2010.
- [13] Norman, G.J., Norris, C., Gollan. J. et.al, "Current Emotion Research in Psychophysiology : The Neurobiology of Evaluative Bivalence, *Emotion Review*, 3, 3349-3359, 2011.
- [14] K. Ulfa, Ragil J., et.al, "Sentiment Analysis of Cyberbullying on Twitter Using SentiStrength", *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, Vol.3, No.1, pp. 21-27, p. 2614-3372, March 2020.