

CLASSIFICATION OF CUSTOMERS EMOTION USING NAÏVE BAYES CLASSIFIER

(Case Study: Natasha Skin Care)

Affah Nurlaila

Program Studi Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
nurlailaaffah@gmail.com

Wiranto

Program Studi Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
wiranto@staff.uns.ac.id

Ristu Saptono

Program Studi Informatika
Universitas Sebelas Maret
Jl. Ir. Sutami No. 36 A Surakarta
ristu.saptono@staff.uns.ac.id

Abstract— Today customers can easily submit their review or opinion of a product or a service from Natasha Skin Care through mentions tweet @NatashaSkinCare. Mentions can be identified emotions of costumers after using a products or a services of Natasha Skin Care. This research proposes to classify emotions according to Ekman that is joy, surprise, anger, fear, sad, and disgust by using Naïve Bayes Classifier. Naïve Bayes Classifier is chosen because of its advantages that is simple, fast, and high accuracy. The dataset in this study amounted to 19,253 with the division for each class is 804 joy, 43 surprise, 154 anger, 61 fear, 287 sad, 167 disgust, and 17736 no-emotions. The results show that the Naïve Bayes Classifier method has a good performance to classify the emotions Natasha Skin Care customers through twitter. The average accuracy rate on the no-emotions class dataset is 80.19%. The average of the emotional classification without involving the no-emotions class shows the highest recall value in the joy class of 92.21%. The highest precision value in the surprise class was 97.77% and the highest F1-Measure was in the joy class of 89.14%. The mean on the dataset with the no-emotions class is 88.58%. Although the accuracy of the dataset with the no-emotions class is higher, the precision and recall value is very low, which is 0%. After using the ROS resampling algorithm, the mean values of precision, recall, and F1-Measure are highest in the no-emotions class. Precision value is 96.64%, recall value is 76.36%, and F1-Measure value is 85.93%.

Keyword: emotional classification, naïve bayes classifier, resampling

1. PENDAHULUAN

Perkembangan industri kecantikan di Indonesia saat ini berkembang semakin pesat. Potensi tersebut muncul karena penduduk Indonesia yang didominasi

oleh perempuan ingin selalu terlihat cantik. Tak hanya perempuan, saat ini para lelaki mulai turut memperhatikan secara keseluruhan penampilan mereka setiap saat. Kondisi tersebut akhirnya dimanfaatkan oleh para produsen klinik kecantikan di Indonesia untuk mengembangkan bisnis mereka.

Klinik kecantikan termasuk dalam usaha dengan tujuan memuaskan pelanggan atau konsumen. Kepuasan pelanggan dapat dilihat dari komentar, opini, atau *review* yang diberikan setelah menggunakan jasa dan/atau produk dari klinik kecantikan. Dewasa ini, komentar, opini, atau *review* yang diberikan pelanggan dapat ditemukan melalui berbagai sosial media.

Salah satu sosial media yang dapat digunakan untuk memberikan komentar, opini, atau *review* adalah *twitter*. Banyak pengguna *twitter* yang secara tidak sadar memberikan informasi tentang kepribadian, emosi, dan juga perasaan yang sedang mereka alami [1]. Emosi merupakan salah satu faktor penting dalam interaksi manusia komputer. Kemampuan mendeteksi dan menanggapi emosi dapat meningkatkan kecerdasan sistem komputer [2].

Sebagai industri kecantikan yang ternama, Natasha Skin Care menggunakan *twitter* sebagai salah satu media untuk dapat berkomunikasi dengan konsumen. Hal ini dapat dilakukan konsumen dengan melakukan *mentions twitter @NatashaSkinCare*. Namun demikian, salah satu kelemahan penyampaian komentar, opini, atau *review* melalui *twitter* adalah *mentions* berbentuk teks digital tidak terstruktur.

Dalam menangani masalah teks tidak terstruktur dapat dilakukan dengan metode analisis *text mining*. *Text mining* merupakan teknik yang digunakan untuk menangani permasalahan klasifikasi, *clustering*,

information extraction, dan *information retrieval*. Selain menggunakan *text mining*, dalam penelitian ini juga menggunakan algoritma *Naïve Bayes Classifier*. *Naïve Bayes Classifier* merupakan metode klasifikasi yang dapat memprediksi probabilitas sebuah *class*, sehingga dapat menghasilkan keputusan berdasarkan data pembelajaran [3].

Penelitian untuk melihat performa *Naïve Bayes Classifier* dalam melakukan klasifikasi emosi telah dilakukan oleh Kaur, J dan Saini, J.R [4]. Dalam penelitian tersebut, didapatkan hasil bahwa metode *Naïve Bayes Classifier* merupakan metode klasifikasi yang paling baik untuk jenis teks dengan bahasa *informal*. Pada penelitian ini akan dilakukan klasifikasi emosi dari konsumen Nataska *Skin Care* yang berasal dari *mentions twitter* dengan menggunakan *text mining* dan *Naïve Bayes Classifier*. Dengan adanya klasifikasi emosi diharapkan dapat memberikan pemahaman kepada pihak Nataska *Skin Care* untuk meningkatkan kualitas pelayanan. Selain itu, klasifikasi emosi dapat juga digunakan untuk calon konsumen dalam pengambilan keputusan.

2. LANDASAN TEORI

2.1 Model Emosi

Dalam Kamus Besar Bahasa Indonesia, emosi didefinisikan sebagai keadaan dan reaksi psikologis dan fisiologis (seperti kegembiraan, kesedihan, keharuan, kecintaan). Model yang paling populer untuk mewakili emosi adalah model kategori dan model dimensi [5]. Model kategori mengasumsikan bahwa ada jumlah emosi dasar yang terbatas dan diskrit, dimana masing-masing emosi tersebut melayani tujuan tertentu. Sedangkan model dimensi mewakili emosi dalam pendekatan dimensi. Dalam pendekatan ini, model dimensi mengasumsikan bahwa ruang emosional dibuat dan setiap emosi terletak pada ruang ini. Sebuah model kategori yang sangat populer dan banyak digunakan adalah emosi model Ekman yang menentukan enam emosi dasar manusia, yaitu marah (*anger*), jijik (*disgust*), takut (*fear*), kebahagiaan (*joy*), kesedihan (*sadness*), kejutan (*surprise*).

2.2 Text Mining

Text mining adalah proses mendapatkan pengetahuan yang intensif dimana pengguna berinteraksi dengan kumpulan dokumen dengan

menggunakan perangkat analisis tertentu. Jika dianalogikan dengan data mining, *text mining* mengekstrak informasi yang berguna dari berbagai sumber data melalui identifikasi dan eksplorasi pola-pola yang menarik. Sumber datanya adalah kumpulan dokumen, dan pola-pola yang menarik yang ditemukan tidak diformalisasi ke dalam *record* di *database* tetapi dalam data tekstual yang tidak terstruktur [6]. Penerapan *text mining* dapat dilakukan pada aplikasi media online, analisis sentimen, aplikasi akademik, *filter email spam*, dan lain sebagainya.

2.3 Text Preprocessing

Tahap *text preprocessing* adalah tahap awal dari *text mining*. Pada tahapan *preprocessing* dilakukan beberapa proses, yaitu:

a. Proses *Tokenizing*

Tahapan ini memisahkan deretan kata atau *tweet* menjadi token. Selain itu juga menghilangkan karakter-karakter tertentu (seperti tanda baca, karakter, angka, *tag HTML*, *link URL*, *username*, dan lain sebagainya) dan mengubah semua token ke bentuk huruf kecil (*lower case*).

b. Proses *Filtering*

Filtering adalah proses penghapusan *stopwords*. *Stopword* adalah teks yang sering muncul di dalam dokumen tapi tidak mengandung informasi yang signifikan. Dengan dihapusnya *stopwords*, ukuran kosakata menjadi berkurang sehingga hanya kata-kata penting yang terdapat dalam dokumen. Contoh *stopwords*, diantaranya: yaitu, adalah, akhirnya, bagaimanapun, yang, dan.

c. Proses *Stemming*

Stemming merupakan proses pemotongan imbuhan atau pengembalian kata berimbuhan menjadi kata dasar. Tujuan utama *stemming* adalah untuk mengubah susunan kata seperti kata benda, kata sifat, kata kerja, kata keterangan, dan sebagainya kedalam bentuk kata dasar. Secara garis besar algoritma *stemming* dapat diklasifikasikan menjadi tiga kategori yaitu *truncating methods*, *statistical methods*, dan *mixed methods* [7]. Contoh dari *truncating methods* adalah algoritma *Porter Stemmer* yang dalam penerapannya akan menghapus *suffix* dan *prefix* dari sebuah kata.

2.4 Resampling

Resampling merupakan konsep penting dalam statistik inferensial. Hal ini digunakan untuk menarik sejumlah besar sampel dari yang asli untuk mencapai perkiraan distribusi teoritis yang mendasari. Hal ini didasarkan pada perulangan sampel dalam kumpulan data yang sama [8]. *Resampling* ini digunakan sebab berkaitan dengan kesulitan dalam mendapatkan dataset yang besar karena datanya rahasia atau datanya belum lengkap di alam. Ada tiga pendekatan dasar untuk mengatasi masalah ketidakseimbangan kelas, yaitu *oversampling* kelas minoritas (ROS), *undersampling* kelas mayoritas (RUS), atau penggunaan pendekatan *hybrid* didasarkan pada keduanya.

2.5 Naïve Bayes Classifier

Naïve Bayes Classifier atau disebut juga dengan *Multinomial Naïve Bayes* merupakan algoritma *Naïve Bayes* yang menangani data multinomial yang digunakan dalam klasifikasi teks. Data dalam *Multinomial Naïve Bayes* direpresentasikan sebagai jumlah vektor kata.

Dalam *Multinomial Naïve Bayes* pertama dilakukan penghitungan probabilitas kata dalam kelas (*prior*) menggunakan persamaan berikut [9]:

$$P(C_i) = \frac{N(C_i)}{N_{total}}$$

$N(C_i)$ adalah jumlah data latih yang termasuk jenis emosi C_i dan N_{total} adalah jumlah seluruh data latih. Tahap pelatihan juga akan menghitung nilai probabilitas dari setiap kata w_k untuk setiap jenis emosi C_i menggunakan persamaan *likelihood* berikut:

$$P(w_k|C_i) = \frac{n_k+1}{n+v}$$

dimana n_k adalah jumlah kemunculan kata w_k pada jenis emosi C_i , n adalah jumlah kata pada tiap jenis emosi dan v adalah jumlah kosakata pada data latih.

Selanjutnya jenis emosi ditentukan berdasarkan nilai probabilitas *posterior* terbesar. Nilai *posterior* maksimal dihitung menggunakan persamaan berikut [9]:

$$C_{MAP} = \operatorname{argmax}_{c_i \in C} P(C_i) \prod_{m=1}^n P(w_k|C_i)$$

dimana $P(C_i)$ menunjukkan nilai probabilitas setiap jenis emosi dan dihitung menggunakan *prior*. Sedangkan $P(w_k|C_i)$ merupakan nilai probabilitas setiap kata pada setiap jenis emosi. Nilai $P(w_k|C_i)$ dihitung menggunakan *likelihood*.

Persamaan *posterior* melakukan perkalian nilai probabilitas yang bertipe *float*. Kondisi tersebut dapat mengakibatkan terjadi *floating-point underflow*. Oleh karena itu persamaan *posterior* diubah menjadi operasi penjumlahan sebagai berikut:

$$C_{MAP} = \operatorname{argmax}_{c_i \in C} \log P(C_i) \sum_m \log P(w_k|C_i)$$

2.6 Confusion Matrix

Confusion matrix merupakan matriks yang menampilkan prediksi klasifikasi dan klasifikasi yang aktual. Unjuk kerja dari model deteksi dapat diukur menggunakan tabel *confusion matrix* untuk mendapatkan nilai *precision*, *recall*, dan *F1-measure* [10]. *Confusion matrix* berukuran $n \times n$ dapat dilihat pada Tabel 1.

Tabel 1 *Confusion Matrix* berukuran $n \times n$

		Predicted-Class				Total
		Kelas 1	Kelas 2	..	Kelas n	
Actual-Class	Kelas 1	<i>True-Positive</i>	<i>Error</i>	..	<i>Error</i>	Total Kelas-1
	Kelas 2	<i>Error</i>	<i>True-Positive</i>	..	<i>Error</i>	Total Kelas-2

	Kelas n	<i>Error</i>	<i>Error</i>	..	<i>True-Positive</i>	Total Kelas-n
		Prediksi Kelas-1	Prediksi Kelas-2	..	Prediksi Kelas-n	

Confusion matrix dapat digunakan untuk mencari nilai dari *precision*, *recall*, *accuracy*, dan *F1-measure* dengan persamaan sebagai berikut:

$$Precision = \frac{TP(Kelas-i)}{Prediksi(Kelas-i)} \times 100\%$$

$$Recall = \frac{TP(Kelas-i)}{Total(Kelas-i)} \times 100\%$$

$$Accuracy =$$

$$\frac{TP(Kelas-1)+TP(Kelas-2)+\dots+TP(Kelas-n)}{Total(Kelas-1)+Total(Kelas-2)+\dots+Total(Kelas-n)} \times 100\%$$

$$F1-measure = (2 \times \frac{precision \times recall}{precision+recall}) \times 100\%$$

3. METODOLOGI

Tahapan metodologi penelitian yang dilakukan adalah:

3.1 Pengumpulan Data

Data yang digunakan adalah *tweet* yang dilakukan oleh konsumen *Natasha Skin Care* yang diketahui dari *mention* pada akun *twitter* @NatashaSkinCare. Pengumpulan data *tweet* dilakukan pada tanggal 31 Mei 2014 sampai dengan 11 Agustus 2017.

3.2 Text Preprocessing

Pada tahapan *text preprocessing* dilakukan beberapa proses, yaitu:

a. Case-folding

Pada tahap ini, seluruh data akan diubah menjadi huruf kecil (*lowercase*).

b. Penghilangan *username* dan *hashtag*

c. Penghilangan karakter selain huruf

d. Tokenizing

e. Filtering

3.3 Penerapan *Naïve Bayes Classifier*

Langkah-langkah yang dilakukan dalam proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier* adalah sebagai berikut:

1. Menghitung nilai *prior* setiap klasifikasi
 2. Menghitung nilai *likelihood* setiap *term* pada *tweet* untuk setiap klasifikasi
 3. Menghitung nilai *posterior* setiap *tweet* untuk setiap klasifikasi
 4. Menentukan klasifikasi emosi berdasarkan nilai *posterior* tertinggi
- ### 3.4 Evaluasi

Tahapan evaluasi merupakan tahap sistem akan diuji dengan menggunakan *confusion matrix* untuk mengetahui nilai *precision*, *recall*, *accuracy*, dan *F1-measure*.

4. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini berasal dari proses *crawling mention @NatashaSkinCare*. Data diambil dari tanggal 31 Mei 2014 sampai dengan 11 Agustus 2017 dengan jumlah 39.201 data. Setelah penghapusan *tweet* dari @NatashaSkinCare dan duplikasi *tweet* maka data yang digunakan pada penelitian ini sejumlah 19.253 data. Dari data yang sudah didapatkan tersebut, kemudian dikelompokkan secara manual kedalam tujuh kelas emosi yaitu *joy*, *surprise*, *anger*, *fear*, *sad*, *disgust*, dan *noemotion*. Pengelompokkan tujuh kelas emosi dapat dilihat pada Tabel 2.

Tabel 2 Data Penelitian

Kelas Emosi	Jumlah Tweet
<i>Noemotion</i>	17736
<i>Joy</i>	804
<i>Surprise</i>	43
<i>Fear</i>	61
<i>Sad</i>	288
<i>Anger</i>	154
<i>Disgust</i>	167

Joy merupakan model emosi berdasarkan perasaan senang, puas, dan sejenisnya terhadap produk atau jasa dari *Natasha Skin Care*. *Surprise* adalah model emosi yang dirasakan konsumen ketika mendapatkan suatu kejutan atau hal yang tidak biasanya. Sedangkan *Anger* adalah model emosi konsumen tatkala merasa marah atau kecewa terhadap produk ataupun pelayanan yang diberikan. *Fear* merupakan emosi konsumen saat merasa takut berkaitan dengan produk atau jasa, baik disebabkan faktor internal atau eksternal. *Sad* adalah emosi sedih konsumen setelah penggunaan produk atau jasa dan juga perasaan kangen atau rindu terhadap *Natasha Skin Care*. *Disgust* mewakili emosi konsumen yang merasa alergi, panas, perih atau sejenisnya setelah penggunaan produk. Selain itu, *disgust* juga dapat digunakan ketika konsumen merasa jijik terhadap pelayanan atau tempat *skin care*.

Pada proses *labelling* data, terdapat kelas emosi yang tidak berkaitan dengan kelas emosi yang dipaparkan Ekman. Kelas emosi yang tidak berkaitan tersebut digolongkan sebagai kelas *noemotion*. *Noemotion* banyak dijumpai pada *tweet* bertipe pertanyaan dan *tagging* tempat.

Selanjutnya dilakukan *text preprocessing* terhadap data tersebut. Setelah proses *text preprocessing* selesai, dilakukan proses klasifikasi terhadap dokumen dengan *Naïve Bayes Classifier*. Pada penelitian ini dilakukan dua tahap yang berbeda. Tahap pertama klasifikasi emosi menggunakan data tanpa *noemotion*. Sedangkan pada tahap kedua, klasifikasi emosi menggunakan data dengan *noemotion*. Hal ini dilakukan untuk mengetahui adakah pengaruh kelas *noemotion* pada data.

Dengan menggunakan perbandingan data *training* dan data *testing* 50:50, pengujian tanpa data *noemotion* menunjukkan akurasi yang cukup baik yaitu sebesar 76,77%. Sedangkan nilai *precision* terendah terdapat pada kelas *Sad* yaitu 64,70%. Nilai *precision* yang rendah diperoleh karena dari 119 *tweet* yang dideteksi sebagai kelas *Sad*, ternyata hanya 77 yang merupakan kelas *Sad*. Sedangkan untuk keseluruhan, model mampu mendeteksi 51 *anger*, 63 *disgust*, 369 *joy*, 7 *fear*, 77 *sad*, dan 88 *surprise*.

Rata-rata nilai *precision* tertinggi terdapat pada kelas *surprise* yaitu sebesar 97,77%. Nilai *recall* tertinggi adalah 92,21% pada kelas *joy*. Sedangkan pada rata-rata hasil *F1-Measure* tertinggi terdapat

pada kelas *joy* dengan nilai sebesar 89,14%. Pada kelas *fear*, *F1-Measure* bernilai tak terhingga dikarenakan pada proses data uji ke-4 dan ke-5, nilai *precision* dan *recall* bernilai 0 (kelas *fear* tidak mampu diklasifikasikan sebagai *fear*).

Sedangkan pada data dengan *noemotion* terjadi ketidakseimbangan kelas. Hal ini disebabkan kelas *Noemotion* mendominasi kelas lainnya. Akurasi yang diperoleh pada data dengan *noemotion* menunjukkan hasil yang lebih baik daripada percobaan sebelumnya. Namun demikian, nilai *precision* dan *recall* sangatlah rendah. Selain itu, terdapat dua kelas yang tidak dapat diklasifikasikan oleh sistem yaitu *fear* dan *surprise*.

Untuk mengatasi ketidakseimbangan kelas, dilakukan proses *resampling* dengan algoritma *oversampling* kelas minoritas (ROS). Proses pada algoritma ROS yaitu melakukan pengulangan data pada kelas minoritas hingga sama dengan jumlah kelas mayoritas pada data latih. Sehingga nilai *anger*, *disgust*, *joy*, *fear*, *sad*, dan *surprise* pada data latih yaitu sebagai berikut:

<i>Joy</i>	:	<i>Surprise</i>	:	<i>Anger</i>	:	<i>Fear</i>	:	<i>Sad</i>	:	<i>Disgust</i>	:	<i>No-emotions</i>
402	:	21	:	77	:	30	:	143	:	83	:	8868



<i>Joy</i>	:	<i>Surprise</i>	:	<i>Anger</i>	:	<i>Fear</i>	:	<i>Sad</i>	:	<i>Disgust</i>	:	<i>No-emotions</i>
8466	:	8847	:	8791	:	8838	:	8725	:	8785	:	8868

Setelah menggunakan algoritma *resampling* ROS nilai *precision*, *recall*, dan *F1-measure* menjadi lebih baik daripada percobaan sebelumnya. Kelas emosi yang sebelumnya tidak dapat diklasifikasikan, akhirnya mampu diklasifikasikan dengan nilai *precision* kelas *Fear* adalah 60% dan nilai *precision* kelas *Surprise* adalah 7,75%. Sedangkan untuk *recall* kelas *Fear* adalah 10% dan *recall* kelas *Surprise* adalah 40,9%. Kedua kelas tersebut mampu diklasifikasikan karena adanya penambahan data pada data *training* sehingga sistem mampu mengklasifikasikan dengan baik. Pada klasifikasi emosi dengan *noemotion*, rata-rata nilai evaluasi *precision*, *recall*, dan *F1-Measure* tertinggi terdapat pada kelas *Noemotion*. Sedangkan untuk rata-rata nilai *precision* terkecil terdapat pada kelas *Surprise* dengan nilai 6,88%. Rata-rata nilai *recall* terkecil adalah pada kelas *Fear* sebesar 18,11% dan rata-rata nilai *F1-Measure* terkecil terdapat pada kelas *Sad* dengan nilai 11,92%.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa metode *Naïve Bayes Classifier* tidak selalu menghasilkan kinerja yang baik dalam melakukan klasifikasi emosi konsumen. Hal ini dapat terjadi pada kasus ketidakseimbangan kelas yang mengakibatkan adanya dua kelas yang tidak dapat diklasifikasikan, yaitu *fear* dan *surprise*.

Pada proses pengujian, rata-rata tingkat *accuracy* pada *dataset* tanpa melibatkan kelas *no-emotions* adalah 80,19%. Rata-rata pada klasifikasi emosi tanpa melibatkan kelas *no-emotions* menunjukkan nilai *recall* tertinggi pada kelas *joy* yaitu sebesar 92,21%. Nilai *precision* tertinggi pada kelas *surprise* yaitu sebesar 97,77% dan nilai *F1-Measure* tertinggi terdapat pada kelas *joy* sebesar 89,14%. Sedangkan rata-rata pada *dataset* dengan kelas *no-emotions* adalah 88,58%. Walaupun *accuracy* pada *dataset* dengan kelas *no-emotions* lebih tinggi, tetapi nilai *precision* dan *recall*-nya sangatlah rendah yaitu mencapai nilai 0%.

Untuk mengatasi ketidakseimbangan kelas tersebut digunakan algoritma *resampling* ROS (*oversampling* kelas minoritas). Algoritma *resampling* ROS dapat meningkatkan nilai *recall*, *precision*, dan *F1-measure*. Setelah menggunakan algoritma *resampling* ROS, rata-rata nilai *precision*, *recall*, dan *F1-Measure* tertinggi terdapat pada kelas *no-emotions*. Yaitu *precision* bernilai 96,64%, *recall* bernilai 76,36%, dan *F1-Measure* bernilai 85,93%.

Untuk perbaikan pada penelitian selanjutnya, saran yang dapat dipertimbangkan adalah sebagai berikut:

1. Model emosi yang digunakan tidak hanya berdasar pada model emosi *Ekman*, tetapi juga model emosi lainnya. Sebagai contoh model Emosi *Russell* yang merupakan model emosi dimensional dengan jenis emosi yang lebih banyak.
2. Dapat menggunakan *feature selection* untuk mengurangi jumlah kata (*term*) pada *matrix*.
3. Pada pengklasifikasian kelas *no-emotions* dan kelas emosi (*joy*, *surprise*, *disgust*, *anger*, *fear*, *sad*) juga menggunakan metode *Naïve Bayes Classifier*.
4. Terdapat peta emosi berdasarkan lokasi Natasha *Skin Care* pada tiap daerahnya.

6. DAFTAR PUSTAKA

- [1] Qiu, L., Lin, H., Ramsay, J., dan Yang, F., "You are what you tweet: Personality expression and perception on Twitter," *Science Direct*, pp. 710-718, 2012.
- [2] Tao, J. & Tan, T., "Affective Computing: A Review," *Affective Computing and Intelligent Interaction*, pp. 981-995, 2005.
- [3] Basuki, A, Metode Bayes, Surabaya: ITS, 2006.
- [4] Kaur, Jasleen; R. Saini, Jatinderkumar, "Emotion Detection and Sentiment Analysis in Text Corpus: A Differential Study with Informal and Formal Writing Styles," *International Journal of Computer Applications*, vol. 101, no. 9, 2014.
- [5] Perikos, I; Hatzilygeroudis, I, "Recognizing emotions in text using ensemble of classifiers," *Engineering Applications of Artificial Intelligence*, 2016.
- [6] R. Feldman, J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2007.
- [7] A. Jivani, "A Comparative Study of Stemming Algorithms," *Int. J. Comp. Tech. Appl.*, p. 1930–1938, 2011.
- [8] C. H. Yu, "Resampling methods: Concepts, Applications, and Justification," *Practical Assessment, Research & Evaluation*, vol. 8, 2003.
- [9] T. Mitsa, Temporal Data Mining, New York: CRC Press, 2010.
- [10] B. & P. V. Binali, "Emotion Detection State of the Art," *Proceedings of the CUBE International Information Technology Conference (CUBE '12)*, pp. 501-507, 2012.