

# PENERAPAN METODE NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI EMOSI PADA KOMENTAR MEDIA SOSIAL

Erfian Junianto<sup>1</sup>, Rizal Rachman<sup>2</sup>

<sup>1</sup>Universitas Adhirajasa Reswara Sanjaya  
e-mail: erfian.ejn@ars.ac.id

<sup>2</sup>STMIK Nusa Mandiri  
e-mail: rizal.rzc@bsi.ac.id

## Abstrak

Keadaan psikologis yang dimiliki setiap orang diantaranya adalah emosi. Keadaan tersebut juga digunakan untuk membentuk hubungan atau menyampaikan pesan antara penulis dengan pembaca. Penetrasi internet yang semakin tinggi membuat gaya hidup masyarakat semakin berubah. Gaya hidup semakin terpengaruh oleh media sosial, sehingga membuat penyampaian informasi dan gagasan menjadi sangat bebas. Informasi dan gagasan tersebut selalu mengandung emosi dari penulisnya. Baik emosi marah, takut, bahagia, ataupun sedih. Pada penelitian ini akan diterapkan model *text mining* untuk melakukan deteksi melalui proses klasifikasi. Metode yang digunakan adalah naive bayes classifier. Sedangkan dataset yang digunakan sudah diberikan kelas oleh peneliti sebelumnya menggunakan best-worst scaling. Tantangan dari dataset ini adalah melakukan dari kalimat atau teks ke dalam 4 kategori emosi yaitu anger, fear, joy, dan sadness. Dari beberapa penelitian sebelumnya metode naive bayes classifier mampu menghasilkan akurasi yang cukup tinggi. Dengan diterapkan pada dataset emosi yang mempunyai karakter berbeda dari dataset kebanyakan, metode tersebut masih mampu memberikan kinerja yang baik dalam klasifikasi.

**Kata Kunci:** Text Mining, Klasifikasi, Deteksi Emosi, Naïve Bayes Classifier

## Abstract

*The psychological state that everyone has is emotional. The situation is also used to form relationships or convey messages between the writer and the reader. The higher internet penetration makes people's lifestyles increasingly change. Lifestyle is increasingly influenced by social media, thus making the delivery of information and ideas very freely. The information and ideas always contain emotions from the author. Both the emotions of anger, fear, happiness, or sadness. In this research a text mining model will be applied to detect through the classification process. The method used is Naive Bayes Classifier. While the dataset used has been given a class by previous researchers using the best-worst scaling. The challenge of this dataset is to carry out sentences or text into 4 categories of emotions, namely anger, fear, joy, and sadness. From several previous studies the Naive Bayes classifier method is capable of producing quite high accuracy. By applying to emotional datasets that have different characters from most datasets, the method is still able to provide good performance in classification.*

**Keywords:** Text Mining, Classification, Emotion Detection, Naive Bayes Classifier

## 1. Pendahuluan

Keadaan psikologis yang menjembatani hubungan antara penulis dengan pembaca, serta memperlihatkan respon seseorang terhadap sesuatu antara tujuan dengan peristiwa di dunia nyata

disebut dengan emosi (Jenkins, Oatley, & Keltner, 2013). Peristiwa meningkatnya pengguna internet membuat penyampaian informasi dengan sangat cepat dan sangat luas. Data survei pengguna internet di Indonesia sendiri pada 2012 yaitu 63 juta

jiwa, tahun 2017 meningkat pesat menjadi 143,6 juta jiwa (Izza, 2017). Layanan yang paling sering diakses adalah sosial media. Sebanyak 87,13% dari pengguna internet di Indonesia mengakses akun media sosial, angka ini mengalahkan penggunaan internet untuk mengakses search engine yang jumlahnya hanya sebesar 74,84% (Izza, 2017).

Media sosial mampu mempengaruhi dan mengubah gaya hidup seseorang (Chukwuere, 2017). Berdasarkan data dari Digital Information World, pengguna facebook menduduki peringkat pertama. Sedangkan twitter masih menduduki peringkat 10 besar (Mehmood, 2018). Perbedaan mendasar terletak pada aturan pembuatan konten, dimana twitter hanya dibatasi sebanyak 140 karakter saja, sedangkan facebook tidak. Konten yang singkat memaksa penulis untuk membuatnya mengandung semua perasaan penulis.

Pengembangan text mining dari dataset twitter adalah sentiment analysis (Ghosh et al., 2015; Rosenthal, Farra, & Nakov, 2017; Rosenthal et al., 2015; Saif & Fernandez, 2016; Santos & Gattit, 2014; Severyn & Moschitti, 2015; Zimbra, Ghiassi, & Lee, 2016). Dalam penelitian tersebut, dipelajari kasus bentuk opini positif dan negatif yang di ekstrak dari berbagai bahasa di dalam tweet. Pengembangan lebih lanjut dari penelitian tersebut adalah pengelompokan tweet dari segi nilai emosi yang terkandung di dalamnya. Didasarkan pada banyaknya data opini di dalam media sosial seperti Facebook, Twitter, Instagram, dan media sosial lainnya, maka data opini tersebut dapat diambil sebagai dataset untuk kemudian dilakukan deteksi emosi dari setiap opininya, dalam hal ini pada setiap tweet. Setiap opini dapat mengekspresikan perasaan seperti persahabatan, dukungan sosial, kemarahan, kebahagiaan, dan lain-lain (Akaichi, Dhouioui, & Lopez-Huertas Perez, 2013).

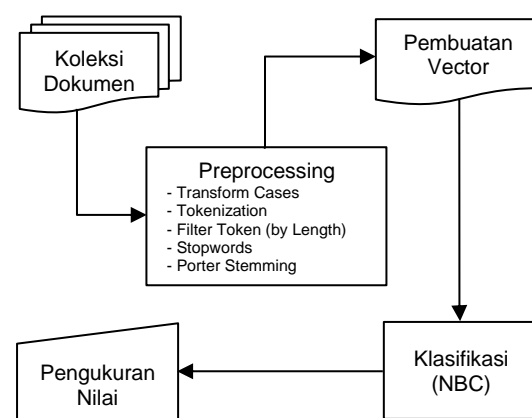
Penelitian sebelumnya banyak yang menekankan pada 2 kategori analisa, yaitu positif atau negatif (Hamouda & Akaichi, 2013; Soelistio & Surendra, 2015; Wakade, Shekar, Liszka, & Chan, 2012), puas atau tidak puas (Yee Liau & Pei Tan, 2014), dan lain sebagainya. Hal ini masuk dalam

kategori sentiment analysis. Perkembangan dari penelitian sebelumnya adalah emotion detection, yang mana akan di ekstrak informasi emosi pada sebuah teks berdasarkan struktur kata dan kalimat. Beberapa penelitian sudah melakukan hal tersebut dan terdapat beberapa kategori emosi yang di deteksi (Shen, Li, Zheng, Ren, & Cheng, 2009). Kategori tersebut antara lain marah, menjijikkan, sedih, senang, mengherankan, dan takut (Yam, 2015). Pada penelitian kali ini digunakan dataset yang berasal dari WASSA-2017 (Mohammad & Bravo-Marquez, 2017b).

Dataset tersebut terdiri dari 3 kelompok yaitu training set, development set, dan testing set. Masing-masing kelompok terdiri dari 4 kategori yaitu anger, fear, joy, dan sadness. Dataset tersebut merupakan data yang diambil dari twitter, dan sebelumnya telah di anotasi intensitasnya menggunakan best-worst scalling secara manual oleh peneliti sebelumnya (Mohammad & Bravo-Marquez, 2017a). Pada penelitian ini akan dilakukan klasifikasi dari dataset tersebut, menggunakan Naïve Bayes Classifier (NBC).

## 2. Metode Penelitian

Dalam pengembangan model konseptual pada penelitian ini akan mengadopsi model yang digunakan oleh peneliti sebelumnya (Junianto & Riana, 2017). Beberapa tahapan seperti Document Collection, Preprocessing, Vector Creation, Optimize Weight, Classification, dan Measurement Values akan diterapkan.



Gambar 1. Metode Penelitian

a. Koleksi Dokumen (*Document Collection*)

Tahap awal penelitian yaitu mengumpulkan dokumen, dimana dokumen tersebut berasal dari WASSA-2017 (Mohammad & Bravo-Marquez, 2017b). Dokumen dataset dibagi dalam 4 kategori, yaitu anger, fear, joy, dan sadness. Sebaran dokumen terlihat dalam Table 1, yang dibagi dalam dokumen training, testing dan development. Namun, yang akan digunakan dalam penelitian ini hanya dokumen training dan testing saja.

Tabel 1. Dataset Emosi Twitter

Emotion	Train	Dev.	Test	All
Anger	857	84	760	1701
Fear	1147	110	995	2252
Joy	823	74	714	1611
sadness	786	74	673	1533
<b>Total</b>	<b>3613</b>	<b>342</b>	<b>3142</b>	<b>7097</b>

Sumber: (Mohammad & Bravo-Marquez, 2017b)

b. *Preprocessing*

Pada tahap ini dilakukan proses untuk mempersiapkan dokumen sebelum masuk ke tahap utama. Tahap preprocessing merupakan titik awal dan tahap yang penting dalam klasifikasi (Durairaj & Ramasamy, 2016). Berikut beberapa tahap preprocessing yang digunakan.

1) *Transform to Lower Cases*

Proses pertama pada tahap preprocessing adalah mengubah semua huruf menjadi lower cases. Hal ini dilakukan agar semua data yang masuk tahap berikutnya menjadi sama dan standar. Sebagai contoh kata "Mathematics", "MaThemaTics", "MatheMatics", dan lain seterusnya akan di transformasi ke dalam kata "mathematics".

2) *Tokenization*

Tahap berikutnya adalah memecah kumpulan kalimat menjadi kata, sekaligus menghilangkan tanda baca. Sehingga akan terbentuk kumpulan kata yang unik dan memuiki arti. Misalnya terdapat kalimat sebagai berikut:

a @firstbsa not turning up? why am i not surprised. late for work again! #fuming

Diubah menjadi:

a firstbsa not turning up why am i not surprised late for work again fuming

3) *Filter Token by Length*

Setelah tahap Transform to Lower Cases dan Tokenization, tahap selanjutnya adalah membuat token dari kumpulan kata. Kata yang kurang memiliki arti atau tidak memenuhi syarat sebagai kata akan dibuang. Tahap ini juga akan mengurangi dimensi dokumen. Misalnya, kata dengan susunan kurang dari 4 huruf dan lebih dari 25 huruf akan dibuang. Sehingga kata yang kurang dari 4 huruf seperti , "a", "an", "my", "rd" atau kata dengan susunan lebih dari 25 huruf seperti "wwwhhhhhaaaaaaaatttttt!!!!!!!", "hhhhhhhhh-hhiiiiiiiiiiiiiiiiiiii" akan dibuang, sehingga akan membentuk token yang lebih bersih dan unik.

4) *Stopwords*

Tahapan ini akan menyempurnakan tahap sebelumnya, yaitu pemilihan token yang benar-benar memiliki hubungan dengan dokumen. Diantara kata yang terdiri dari 3 hingga 30 karakter akan dilakukan seleksi. Walaupun kata-kata tersebut memang menjadi bagian dari sebuah kalimat, namun tidak dapat digunakan sebagai token, karena berupa kata sambung atau keterangan. Misalnya, "the", "this", "that", "these", "those", "her", "his", "its", "my", "our", "their", "your", "all", "few", "many", dan lain-lain.

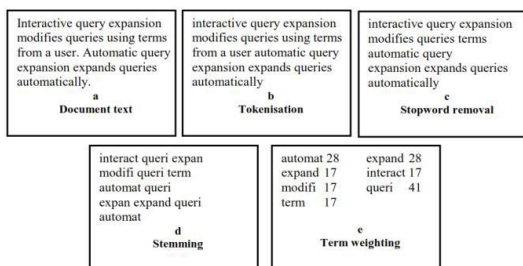
5) *Stemming*

Proses sebelumnya sudah menghasilkan token yang bersih dan unik. Namun, masih ada kata yang sebenarnya memiliki arti yang sama namun berbeda kata. Hal ini dikarenakan pemakain bentuk kata yang berbeda dalam setiap kalimat. Untuk mengatasi hal tersebut pada tahap ini setiap kata akan dikembalikan dalam bentuk akarnya. Proses ini disebut

dengan stemming word. Misalnya kata “disease”, “diseases”, “diseased”, akan diubah menjadi bentuk akar yaitu “disease”. Algoritma stemming yang paling umum dan efektif diterapkan pada bahasa Inggris adalah Porter stemmer (Manning, Raghavan, & Schütze, 2009).

c. Pembuatan Vektor (*Vector Creation*)

Setelah preprocessing selesai, maka tahap selanjutnya adalah mengubah kumpulan kata menjadi vector menggunakan pembobotan TF-IDF (Dadgar, Araghi, & Farahani, 2016; Hakim, Erwin, Eng, Galinium, & Muliady, 2015). Dari vector ini akan dimulai proses optimasi dan klasifikasi. Ilustrasi dari semua tahapan yang juga disebut sebagai indexing, sebagaimana digambarkan pada gambar 2.



Gambar 2. Pengindeksan Dokumen  
Sumber: (Ruthven & Lalmas, 2003)

Pada tahap ini akan dilakukan 2 tahap yaitu term weighting, yang akan menghitung frekuensi kemunculan term dalam dokumen (term frequency), serta total kemunculannya. Kemudian dihitung inverse document frequency (IDF), dan menghasilkan bobot (weight) dengan perkalian antara TF dan IDF. Contoh perhitungan dengan menggunakan TF-IDF sebagai berikut:

Tabel 2. Contoh Koleksi Dokumen

No	Document	Content of Document
1	DC1	the bee sting still suck i feel sick
2	DC2	we were the only ones that were supposed to know that you make me nervous
3	DC3	I thought the holidays could not get any more cheerful, and then I met you
4	DC4	Summer officially ends today

Tabel 3. Contoh Distribusi Term Frequencies

No	Term	DC1	DC2	DC3	DC4	Total (DF)	IDF
1	cheerful	0	0	1	0	1	0,60206
2	could	0	0	1	0	1	0,60206
3	ends	0	0	0	1	1	0,60206
4	feel	1	0	0	0	1	0,60206
5	holidays	0	0	1	0	1	0,60206
6	know	0	1	0	0	1	0,60206
7	make	0	1	0	0	1	0,60206
8	more	0	0	1	0	1	0,60206
9	nervous	0	1	0	0	1	0,60206
10	officially	0	0	0	1	1	0,60206
11	ones	0	1	0	0	1	0,60206
12	only	0	1	0	0	1	0,60206
13	sick	1	0	0	0	1	0,60206
14	still	1	0	0	0	1	0,60206
15	sting	1	0	0	0	1	0,60206
16	suck	1	0	0	0	1	0,60206
17	summer	0	0	0	1	1	0,60206
18	supposed	0	1	0	0	1	0,60206
19	that	0	2	0	0	1	0,60206
20	then	0	0	1	0	1	0,60206
21	thought	0	0	1	0	1	0,60206
22	today	0	0	0	1	1	0,60206
23	were	0	1	0	0	1	0,60206

d. Klasifikasi dengan Naïve Bayes Classifier (NBC)

Setelah dokumen melalui tahap sebelumnya, dihasilkan data yang benar-benar siap untuk diklasifikasikan. Dengan menggunakan algoritma Naïve Bayes Classifier (NBC), pada tahap ini akan dilakukan perhitungan statistik. Dimana akan dihasilkan nilai kemungkinan (probabilitas) dari sebuah dokumen yang dapat menentukan kelasnya. Diharapkan proses sebelumnya dapat mempermudah klasifikasi sekaligus meningkatkan akurasi (Asy'arie & Pribadi, 2009; Hassan, Rafi, & Shaikh, 2011; Junianto & Riana, 2017; Kamruzzaman & Rahman, 2004).

e. Pengukuran Nilai (*Measurement Values*)

Tahap akhir adalah melakukan pengukuran tingkat akurasi dari tahapan klasifikasi. Hasilnya akan disajikan dalam bentuk table confusion matrix yang menunjukkan nilai akurasi.

### 3. Hasil dan Pembahasan

Keluaran yang dihasilkan dapat dianalisa untuk mendapatkan informasi yang berguna. Pada bagian ini akan dijelaskan tentang teknik Performance Evaluation, Simulation, dan pembahasan algoritma yang diusulkan.

#### 3.1. Performance Evaluation

Algoritma klasifikasi menghasilkan keluaran yang dapat diukur. Hasil dari pengukuran tersebut mencerminkan seberapa baik algoritma tersebut bekerja. Confusion Matrix adalah salah satu metode yang dapat digunakan untuk melakukan pengukuran. Pada dasarnya, confusion

matrix terdiri dari informasi perbandingan antara hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi actual (Powers, 2008).

Confusion Matrix merupakan pengukuran yang menggunakan beberapa istilah sebagai perwakilan dari proses klasifikasi. Istilah tersebut yaitu, True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). TN merupakan total dari data negative yang terdeteksi dengan benar, sedangkan untuk FP merupakan kebalikannya. TP adalah hasil dari jumlah data yang positive yang terdeteksi dengan benar. Kebalikannya adalah FN, dimana data positive terdeteksi sebagai data negative (Fawcett, 2006). Nilai TN, FP, TP, dan FN menghasilkan akurasi, presisi, dan recall. Berikut persamaan yang digunakan untuk menghitung akurasi, presisi, dan recall.

$$\text{precision} = (TP/(TP+FP)) \times 100\% \quad (1)$$

$$\text{recall} = (TP/(TP+FN)) \times 100\% \quad (2)$$

$$\text{accuracy} = ((TP+TN)/((TP+FN) + (FP+TN))) \times 100\% \quad (3)$$

### 3.2. Simulation

Implementasi dari model klasifikasi akan dibahas di bagian ini. Mulai dari Document Collection, hingga mendapatkan nilai akurasi.

Tabel 3. Contoh Koleksi Dokumen

No	Kelas	Dokumen
1	anger	Get to work and there's a fire drill. #fire #outthere #inthedark
2		can't believe Achilles killed me!
3	fear	@Gotham3 - Crushes are for kids! Just like trix. \rHow bout some full blown up Love that makes you quiver deep down into your core? Y E S *="
4		I saw someone discourage someone from following their dreams. Just because you want to live in mediocrity doesn't mean someone else should.
5	joy	Thinking about trying some comedy on youtube. Always been fond of it. Time to nut up. #laughter #comedy #maybeoneday #hopefullyfunny #LOL
6		I saw her, whose countenance expressed a mighty rivers were monotonous, and her lively conversation took care of the various
7	sadness	@MxJackMonroe I couldn't get on with it either. Bits started drooping that shouldn't droop. GP said mooncup alone to blame.
8		@Nataliem55 sadly, war has often been the factor that jump starts US economic growth

Tabel 4. Tahap Preprocessing (Transform to Lower Cases)

No	Kelas	Dokumen
1	anger	get to work and there's a fire drill. #fire #outthere #inthedark
2		can't believe achilles killed me!

3	fear	@gotham3 - crushes are for kids! just like trix. \rhow bout some full blown up love that makes you quiver deep down into your core? y e s *="
4		i saw someone discourage someone from following their dreams. just because you want to live in mediocrity doesn't mean someone else should.
5	joy	thinking about trying some comedy on youtube. always been fond of it. time to nut up. #laughter #comedy #maybeoneday #hopefullyfunny #lol
6		i saw her, whose countenance expressed a mighty rivers were monotonous, and her lively conversation took care of the various
7	sadness	@mxjackmonroe i couldn't get on with it either. bits started drooping that shouldn't droop. gp said mooncup alone to blame.
8		@nataliem55 sadly, war has often been the factor that jump starts us economic growth

Tabel 5. Tahap Preprocessing (Tokenization)

No	Kelas	Dokumen
1	anger	get to work and there s a fire drill fire outthere inthedark
2		can t believe achilles killed me
3	fear	gotham crushes are for kids just like trix rhow bout some full blown up love that makes you quiver deep down into your core y e s
4		i saw someone discourage someone from following their dreams just because you want to live in mediocrity doesn t mean someone else should
5	joy	thinking about trying some comedy on youtube always been fond of it time to nut up laughter comedy maybeoneday hopefullyfunny lol
6		i saw her whose countenance expressed a mighty rivers were monotonous and her lively conversation took care of the various
7	sadness	mxjackmonroe i couldn t get on with it either bits started drooping that shouldn t droop gp said mooncup alone to blame
8		nataliem sadly war has often been the factor that jump starts us economic growth

Tabel 6. Tahap Preprocessing (Filter Token by Length)

No	Kelas	Dokumen
1	anger	get work and there fire drill fire outthere inthedark
2		can believe achilles killed
3	fear	gotham crushes are for kids just like trix rhow bout some full blown love that makes you quiver deep down into your core
4		saw someone discourage someone from following their dreams just because you want live mediocrity doesn mean someone else should
5	joy	thinking about trying some comedy youtube always been fond time nut laughter comedy maybeoneday hopefullyfunny lol
6		saw her whose countenance expressed mighty rivers were monotonous and her lively conversation took care the various
7	sadness	mxjackmonroe couldn get with either bits started drooping that shouldn droop said mooncup alone blame
8		nataliem sadly war has often been the factor that jump starts economic growth

Tabel 7. Tahap Preprocessing (Stopwords)

No	Kelas	Dokumen
1	anger	get work fire drill fire outthere inthedark
2		believe achilles killed
3	fear	gotham crushes kids trix rhow bout full blown love makes quiver deep core
4		saw discourage dreams want live mediocrity mean else
5	joy	thinking trying comedy youtube fond time nut laughter comedy maybeoneday hopefullyfunny lol
6		saw countenance expressed mighty rivers monotonous lively conversation took care various
7	sadness	mxjackmonroe get bits started drooping droop said mooncup blame
8		nataliem sadly war factor jump starts economic growth

Tabel 8. Tahap *Preprocessing* (*Stemming*)

No	Kelas	Dokumen
1	anger	get work fire drill fire outther inthedark
2		believ achil kill
3	fear	gotham crush kid trix rhov bout full blown love make quiver deep core
4		saw discourag dream want live mediocr mean els
5	joy	think try comedi youtub fond time nut laughter comedi maybeonedai hopefullyfunni lol
6		saw counten express mighti river monoton live convers took care variou
7	sadness	mjackmonro get bit start droop droop said mooncup blame
8		nataliem sadli war factor jump start econom growth

Hasil tersebut kemudian dihitung frekuensinya untuk mendapatkan vector menggunakan TF-IDF. Dari hasil vector kemudian dilakukan proses pembobotan menggunakan PSO. Pada proses pembobotan akan dipilih dari index inertia weight 0.2 hingga 1.0. Proses utamanya adalah klasifikasi menggunakan algoritma NBC, dimana dilakukan validasi menggunakan cross validation. Hasil akhir yang didapat berupa akurasi.

### 3.3. Diskusi

Proses uji coba dalam penelitian ini akan dilakukan dengan beberapa kondisi. Uji coba pertama menggunakan algoritma NBC tanpa adanya optimasi, hasil akurasi yang didapat yaitu 65,93% (Lihat table 8).

Tabel 8. Confusion Matrix NBC

	true anger	true fear	true joy	true sadness	class precision
pred. anger	637	156	115	172	<b>58.98%</b>
pred. fear	101	835	104	155	<b>69.87%</b>
pred. joy	47	55	522	71	<b>75.11%</b>
pred. sadness	72	101	82	388	<b>60.34%</b>
class recall	<b>74.33%</b>	<b>72.80%</b>	<b>63.43%</b>	<b>49.36%</b>	

## 4. Kesimpulan

Penelitian ini dilakukan untuk mengetahui seberapa baik NBC untuk mengklasifikasikan Emosi dari data yang berbentuk teks. Pada penelitian sebelumnya algoritma tersebut banyak digunakan untuk melakukan klasifikasi, yang menitik beratkan pada kategori umum. Misalnya, positif atau negatif, pengelompokan dokumen berdasarkan isi, seperti berita olahraga atau berita politik. Akurasi yang didapat cukup tinggi mengingat kata yang menjadi

penyusun dokumen ditulis dengan kata kunci yang mencerminkan kelompoknya. Sedangkan pada penelitian ini, proses klasifikasi didasarkan pada emosi.

Tantangan yang muncul adalah bagaimana cara agar sebuah kalimat atau kata dapat digunakan untuk mengetahui emosi seseorang saat menulisnya. Karena dalam satu suku kata dapat mewakili beberapa emosi. Maka dari itu akurasi yang dihasilkan tidak dapat dibandingkan dengan penelitian sebelumnya, yang didasarkan pada kategori umum.

## Referensi

- Akaichi, J., Dhouioui, Z., & Lopez-Huertas Perez, M. J. (2013). Text mining facebook status updates for sentiment classification. *2013 17th International Conference on System Theory, Control and Computing, ICSTCC 2013; Joint Conference of SINTES 2013, SACCS 2013, SIMSIS 2013 - Proceedings*, 640–645. <https://doi.org/10.1109/ICSTCC.2013.6689032>
- Asy'arie, A. D., & Pribadi, A. W. (2009). Automatic news articles classification in Indonesian language by using Naive Bayes Classifier method. *Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services - I1WAS '09*, 658. <https://doi.org/10.1145/1806338.1806463>
- Chukwuere, J. E. (2017). *The impact of social media on social lifestyle : A case study of university female students*. (December).
- Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016). A novel text mining approach based on TF-IDF and support vector machine for news classification. *Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016*, 112–116. <https://doi.org/10.1109/ICETECH.2016.7569223>
- Durairaj, M., & Ramasamy, N. (2016). A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate. *International Journal of Control Theory and Applications*, 9(27), 255–260. Retrieved from

- <https://www.researchgate.net/publication/312225257>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., & Reyes, A. (2015). SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 470–478. <https://doi.org/10.18653/v1/S15-2080>
- Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2015). Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. *Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014*, 1–4. <https://doi.org/10.1109/ICITEED.2014.7007894>
- Hamouda, S. Ben, & Akaichi, J. (2013). Social Networks' Text Mining for Sentiment Classification: The case of Facebook' statuses updates in the "Arabic Spring" Era. *International Journal of Application or Innovation in Engineering & Management*, 2(5), 470–478. Retrieved from <http://www.ijaiem.org/Volume2Issue5/IJAIEM-2013-05-26-063.pdf>
- Hassan, S., Rafi, M., & Shaikh, M. S. (2011). Comparing SVM and Naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment. *Proceedings of the 14th IEEE International Multitopic Conference 2011, INMIC 2011*, 31–34. <https://doi.org/10.1109/INMIC.2011.6151495>
- Izza, J. (2017). Infografis Penetrasi dan Perilaku Pengguna Internet Indonesia Tahun 2017. In *Apjii*. Retrieved from [apjii.or.id](http://apjii.or.id)
- Jenkins, J. M., Oatley, K., & Keltner, D. (2013). *Understanding Emotions, 3rd Edition* (3rd ed.). Retrieved from [http://bcs.wiley.com/he-](http://bcs.wiley.com/he-bcs/Books?action=index&itemId=111814743X&bcsId=7957)
- Junianto, E., & Riana, D. (2017). Penerapan PSO Untuk Seleksi Fitur Pada Klasifikasi Dokumen Berita Menggunakan NBC. *JURNAL INFORMATIKA*, 4(1). Retrieved from [https://s3.amazonaws.com/academia.edu.documents/52950408/INF-1810-4093-2-PB.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1558598523&Signature=8lawHoF6VvmzkdUyFLe007DOhI%3D&response-content-disposition=inline%3Bfilename%3DPenerapan\\_PSO\\_Untuk\\_Seleksi\\_Fitur\\_Pada\\_K.pdf](https://s3.amazonaws.com/academia.edu.documents/52950408/INF-1810-4093-2-PB.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1558598523&Signature=8lawHoF6VvmzkdUyFLe007DOhI%3D&response-content-disposition=inline%3Bfilename%3DPenerapan_PSO_Untuk_Seleksi_Fitur_Pada_K.pdf)
- Kamruzzaman, S. M., & Rahman, C. M. (2004). Text Categorization Using Association Rule and Naïve Bayes Classifier. *Asian Journal of Information Technology (AJST)*. <https://doi.org/10.3923/ajit.2004.657.665>
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). IEEE Photonics Technology Letters information for authors. In *IEEE Photonics Technology Letters* (Vol. 21). <https://doi.org/10.1109/LPT.2009.2020494>
- Mehmood, M. (2018). Essential Insights Into Internet, Social Media, Mobile And Ecommerce Use Around The World / Digital Information World. Retrieved August 29, 2018, from Digital Information World website: <https://www.digitalinformationworld.com/2018/07/global-internet-stats-infographic.html>
- Mohammad, S. M., & Bravo-Marquez, F. (2017a). *Emotion Intensities in Tweets*. <https://doi.org/10.18653/v1/S17-1007>
- Mohammad, S. M., & Bravo-Marquez, F. (2017b). WASSA-2017 Shared Task on Emotion Intensity. *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*. <https://doi.org/10.18653/v1/W17-5205>
- Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach. Learn. Technol.*, 2.
- Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. *Proceedings of the*

- 11th International Workshop on Semantic Evaluations (SemEval-2017), 502–518. Retrieved from <http://aclweb.org/anthology/S/S17/S17-2088.pdf>
- Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S. M., Ritter, A., & Stoyanov, V. (2015). SemEval-2015 Task 10: Sentiment Analysis in Twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 451–463. Retrieved from <http://alt.qcri.org/semeval2015/cdrom/pdf/SemEval078.pdf>
- Ruthven, I., & Lalmas, M. (2003). A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2), S0269888903000638. <https://doi.org/10.1017/S0269888903000638>
- Saif, H., & Fernandez, M. (2016). Contextual Semantics for Sentiment Analysis of Twitter Categories and Subject Descriptors. *Elsevier*, 52, 5–19. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0306457315000242>
- Santos, C. N. Dos, & Gattit, M. B. T.-I. C. on C. L. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. *Aclweb.Org*. Retrieved from <http://www.aclweb.org/anthology/C14-1008>
- Severyn, A., & Moschitti, A. (2015). Twitter Sentiment Analysis with Deep Convolutional Neural Networks. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*, 959–962. <https://doi.org/10.1145/2766462.2767830>
- Shen, Y., Li, S., Zheng, L., Ren, X., & Cheng, X. (2009). Emotion mining research on micro-blog. *Proceedings - 2009 1st IEEE Symposium on Web Society, SWS 2009*, 71–75. <https://doi.org/10.1109/SWS.2009.5271711>
- Soelistio, Y. E., & Surendra, M. R. S. (2015). *Simple Text Mining for Sentiment Analysis of Political Figure Using Naive Bayes Classifier Method*. <https://doi.org/10.12962/p9772338185001.a18>
- Wakade, S., Shekar, C., Liszka, K. J., & Chan, C.-C. (2012). Text mining for sentiment analysis of Twitter data. *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, 1. Retrieved from <http://world-comp.org/p2012/IKE3997.pdf>
- Yam, C. . (2015). Emotion Detection and Recognition from Text Using Deep Learning - Developer Blog. Retrieved August 23, 2018, from Microsoft website: <https://www.microsoft.com/developerblog/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning/>
- Yee Liau, B., & Pei Tan, P. (2014). Gaining customer knowledge in low cost airlines through text mining. *Industrial Management & Data Systems*, 114(9), 1344–1359. <https://doi.org/10.1108/IMDS-07-2014-0225>
- Zimbra, D., Ghiassi, M., & Lee, S. (2016). Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 1930–1938. <https://doi.org/10.1109/HICSS.2016.244>