

Analisis Sentimen Produk Penjualan Shopee Pada Pengguna Twitter Menggunakan Metode K-Means

Hotpentalia Nababan¹, Indra Kelana Jaya², Samuel Manurung³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Jul 21, 2023
Revised, Aug 15, 2023
Accepted, Aug 21, 2023

Keywords:

Twitter,
Shopee,
K-Means,
TF-IDF

ABSTRAK

Gaya hidup dalam memenuhi kebutuhan dengan membeli secara langsung mulai ditinggalkan sebagai pencegahan untuk mengurangi persebaran covid-19 dengan memanfaatkan jasa jual beli online melalui *e-commerce*. Platform *e-commerce* yang paling banyak digunakan oleh masyarakat Indonesia adalah *mobile marketplace* yaitu Shopee. Pengalaman konsumen dalam mereview produk salah satunya untuk produk *skincare* yang saat ini tidak terlepas dalam memenuhi kebutuhan pribadi. SKINTIFIC merupakan brand lokal Indonesia yang sedang viral dengan penjualan yang cukup banyak dan berbagai *offline store* maupun *online store* juga menyediakan produk ini, karena banyaknya peminat konsumen menggunakan *skincare* tersebut. Keluhan dan opini pengguna Shopee ini pun disalurkan melalui berbagai sarana salah satunya di Twitter. Aplikasi yang menjadi objek penelitian adalah Twitter yang merupakan salah satunya sarana penyampaian review keluhan dan opini masyarakat pengguna shopee. Metode yang digunakan adalah *K-Means Clustering*. Pada penelitian ini menggunakan metode TF-IDF untuk pembobotan kata dan metode *K-Means Clustering* untuk pengelompokan sentimen positif dan negatif. Berdasarkan hasil, pengelompokan dibagi menjadi 2 yaitu *Cluster* Positif dan *Cluster* negatif. Dengan persentase jumlah *cluster* positif (C1) 79,5%. Sedangkan persentase jumlah *cluster* negatif (C2) 19,7%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Hotpentalia Nababan,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: lianababan142@gmail.com

1. PENDAHULUAN

Pengalaman konsumen dalam mereview produk salah satunya untuk produk *skincare* yang saat ini tidak terlepas dalam memenuhi kebutuhan pribadi. SKINTIFIC merupakan brand lokal Indonesia yang sedang viral dengan penjualan yang cukup banyak dan berbagai *offline store* maupun *online store* juga menyediakan produk ini, karena banyaknya peminat konsumen menggunakan *skincare* tersebut.

Gaya hidup dalam memenuhi kebutuhan dengan membeli secara langsung mulai ditinggalkan sebagai pencegahan untuk mengurangi persebaran Covid-19 dengan memanfaatkan jasa jual beli *online* melalui *e-commerce*. Keluhan dan opini pengguna Shopee ini pun disalurkan melalui berbagai sarana salah satunya di Twitter.

Twitter merupakan media sosial yang sangat populer. Meski begitu Twitter pernah mengalami masa penurunan aktivitas pengguna semenjak munculnya saingan media sosial sekelas Twitter namun memiliki fokus konten gambar yaitu Instagram. Saat ini Twitter mulai banyak digunakan kembali oleh masyarakat. Hal yang paling menarik pada Twitter adalah update nya informasi yang sangat cepat, Hal itu didukung oleh adanya fitur *Trends for you*. Pada fitur ini akan ditampilkan serangkaian kata yang trend dalam arti banyak yang membuat postingan kata tersebut dengan jarak waktu yang kecil sehingga begitu cepat total jumlah posting dengan kata tersebut. Sehingga pengguna Twitter akan mendapatkan informasi yang sesuai dengan trend yang tampil saat itu.

Sentiment Analysis adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media social. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut.

Analisis sentimen dilakukan untuk mengetahui kecenderungan opini seseorang terhadap sebuah peristiwa atau masalah, apakah cenderung beropini positif atau negatif. Teknik yang digunakan yaitu text mining. Text mining merupakan sebuah teknik yang digunakan untuk mengekstraksi informasi yang berguna dari data teks yang tidak terstruktur. Text mining mengekstraksi kata kunci atau mengekstraksi pendapat dan ulasan analisis text sehingga dapat mendukung untuk memahami pendapat masyarakat dalam data text.

2. METODE PENELITIAN

Analisa metode merupakan langkah-langkah sistematis dalam melakukan proses perhitungan yang direpresentasikan pada kesesuaian data yang didapatkan dari tempat penelitian.

2.1 Metode K-Means

Metode *K-Means Clustering* ini dimulai dengan menentukan terlebih dahulu berapa cluster yang ingin dibuat, selanjutnya menentukan nilai awal untuk masing-masing *cluster*, lalu menghitung jarak dari masing-masing data dengan nilai awal yang telah ditentukan. Setelah jarak dari tiap data dihitung, data dimasukkan dalam cluster terdekat. Berikut ini merupakan langkah-langkah perhitungan metode *K-Means Clustering*.

1. Menghitung jarak centroid

$$d(X_i, Y_j) = \sqrt{[(X_i - Y_j)^2]}$$

Keterangan :

X_i = Data kriteria

Y_j = Centroid pada *cluster* ke-j

2. Menghitung nilai centroid baru

$$C_k = \frac{1}{N_k} \sum d_i$$

Keterangan :

C_k = *Centroid* baru

N_k = Jumlah *cluster*

d_i = Dokumen dalam *cluster*

2.2 Metode TF-IDF

Dalam metode merupakan gabungan dari 2 konsep untuk menghitung bobot istilah. *Term Frequency* adalah jumlah suatu istilah muncul dalam sebuah dokumen. *Inverse Document Frequency* adalah pengukur seberapa penting istilah tersebut. Berikut adalah rumus perhitungannya.

$$W_{d,t} = tf_{d,t} * IDF_t$$

Keterangan :

d= dokumen ke-d

t = kata ke- t dari kata kunci
 W = bobot dokumen ke- d terhadap kata ke- t

$$IDF = \log \left(\frac{D}{df_n} \right)$$

Keterangan:

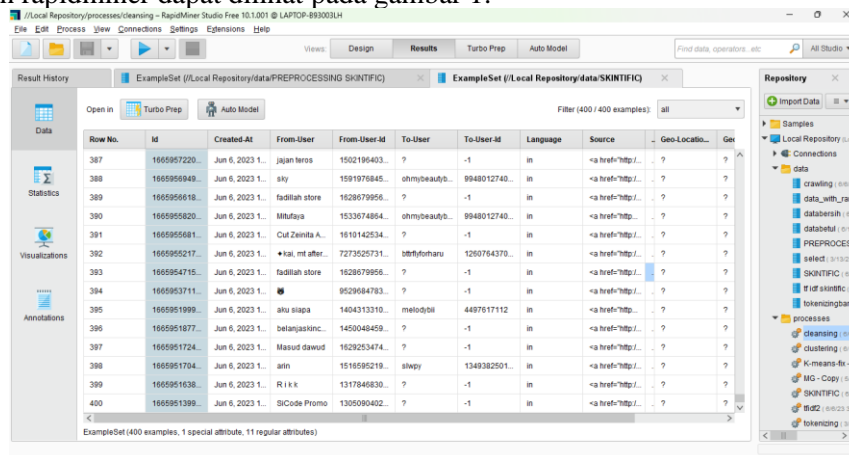
D = total seluruh dokumen
 df = jumlah frekuensi kemunculan kata
 n = data ke- n dari hasil token

3. HASIL DAN PEMBAHASAN

Implementasi analisis sentimen di Twitter dilakukan dengan menggunakan RapidMiner untuk melakukan tahapan *preprocessing*, TF-IDF dan *Clustering* dengan K-Means terhadap data yang sudah ada.

3.1. Pengumpulan data

Data yang berupa *tweet* pengguna twitter dikumpulkan menggunakan aplikasi rapid miner dan disimpan kedalam file data berformat csv. Data yang dikumpulkan dengan proses *crawling* menggunakan rapidminer dapat dilihat pada gambar 1.

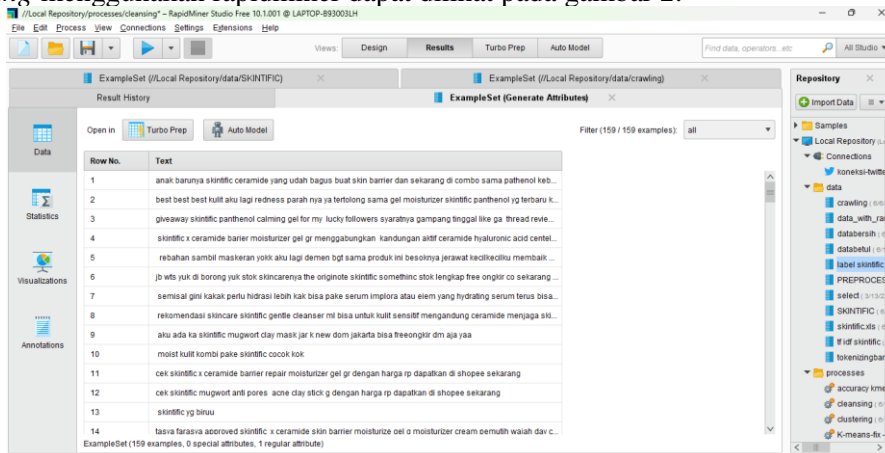


Row No.	Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Geo-Locali
387	1665957220	Jun 6, 2023 1...	jajan teros	1502195403...	?	-1	in	+a href="http://...	?
388	1665956949	Jun 6, 2023 1...	sky	1591976845...	ohmybeauty...	9948012740...	in	+a href="http://...	?
389	1665956616	Jun 6, 2023 1...	fadilah store	1628679956...	?	-1	in	+a href="http://...	?
390	1665955820	Jun 6, 2023 1...	Mitafaja	1533674864...	ohmybeauty...	9948012740...	in	+a href="http://...	?
391	1665955681	Jun 6, 2023 1...	Cut Zeinfa A.	1610142534...	?	-1	in	+a href="http://...	?
392	1665955217	Jun 6, 2023 1...	*kai. mt after.	7273525731...	btbfyforharu	1260764370...	in	+a href="http://...	?
393	1665954715	Jun 6, 2023 1...	fadilah store	1628679956...	?	-1	in	+a href="http://...	?
394	1665953711	Jun 6, 2023 1...	?	952964783...	?	-1	in	+a href="http://...	?
395	1665951999	Jun 6, 2023 1...	aku siapa	1404313310...	melodipi	4497617112	in	+a href="http://...	?
396	1665951877	Jun 6, 2023 1...	belanjakinc...	1450048459...	?	-1	in	+a href="http://...	?
397	1665951724	Jun 6, 2023 1...	Masud dawud	1629253474...	?	-1	in	+a href="http://...	?
398	1665951704	Jun 6, 2023 1...	aini	1516595219...	silwaj	1349382501...	in	+a href="http://...	?
399	1665951638	Jun 6, 2023 1...	R i k k	1317846830...	?	-1	in	+a href="http://...	?
400	1665951399	Jun 6, 2023 1...	SiCode Promo	1305090402...	?	-1	in	+a href="http://...	?

Gambar 1. *Crawling* menggunakan Rapidminer

3.2. Preprocessing

Data yang dimasukkan masih mentah yang masih kotor, sehingga hasil dari tahap ini adalah dokumen berkualitas yang dapat mempermudah proses klasifikasi sentimen. Hasil tahapan *preprocessing* menggunakan rapidminer dapat dilihat pada gambar 2.



Row No.	Text
1	anak barunya skintific ceramide yang udah bagus buat skin barrier dan sekarang di combo sama panthenol keb...
2	best best best kulit aku lagi redness parah nya ya tertolong sama gel moisturizer skintific panthenol yg terbaru k...
3	giveaway skintific panthenol calming gel for my lucky followers syarahnya gampang tinggal like ga thread reve...
4	skintific x ceramide barrier moisturizer gel gr menggabungkan kandungan aktif ceramide hyaluronic acid centel...
5	rebahan sambil maskeran yoiik aku lagi demen bgt sama produk ini besolnya jerawat kecilcilu kembali...
6	jd wta yuk di borong yuk stok skincarenya the originate skintific something stok lengkap free ongkir co sekarang...
7	semisal gini katak perlu hidrasi lebih kak bisa pake serum implora atau elem yang hydrating serum terus bisa...
8	rekomendasi skincare skintific gentle cleanser mi bisa untuk kulit sensitif mengandung ceramide menjaga ski...
9	aku ada ka skintific mugwort clay mask jar k new dom jakarta bisa freeongkir dm aja yaa
10	moist kulit kombi pake skintific cocok loak
11	cek skintific x ceramide barrier repair moisturizer gel gr dengan harga rp dapatkan di shopee sekarang
12	cek skintific mugwort anti pores acne clay stick g dengan harga rp dapatkan di shopee sekarang
13	skintific yg biru
14	tasva farasva approved skintific x ceramide skin barrier moisturizer oel o moisturizer cream oemulth walah dav c...

Gambar 2. Hasil *Preprocessing*

3.3. Pembobotan Data

Perhitungan dengan cara memberikan bobot relasi pada tiap suku kata (*term*) terhadap kata. Kata tunggal tiap kalimat dianggap sebagai dokumen. Pembobotan data pada rapidminer dapat dilihat pada gambar 3.

Row No.	Id	aaaa	abadi	abis	abiss	acid	acne	acnes	ad	adem
1	1666031594.	0	0	0	0	0	0	0	0	0
2	1666032907.	0	0	0	0	0	0	0	0	0,193
3	1666037325.	0	0	0	0	0	0	0	0	0
4	1668410667.	0	0	0	0	0,219	0	0	0	0
5	1668409898.	0	0	0	0	0	0	0	0	0
6	1668409094.	0	0	0	0	0	0	0	0	0
7	1668406040.	0	0	0	0	0	0	0	0	0
8	1668402604.	0	0	0	0	0,189	0	0	0	0
9	1668399949.	0	0	0	0	0	0	0	0	0
10	1668388854.	0	0	0	0	0	0	0	0	0
11	1668384214.	0	0	0	0	0	0	0	0	0
12	1668393974.	0	0	0	0	0	0,241	0	0	0
13	1668393516.	0	0	0	0	0	0	0	0	0
14	1668392001.	0	0	0	0	0	0	0	0	0

Gambar 3. Pembobotan Data

3.4. Pengujian dan Perhitungan metode *K-Means Clustering*

Tahapan ini akan dilakukan proses pengelompokan sentimen positif dan negatif menggunakan metode *K-Means Clustering* menggunakan rapidminer. Dengan $k=2$ yaitu sentimen positif dan negatif dapat dilihat pada gambar 4.

Performance
PerformanceVector:
Avg. within centroid distance: 0,949
Avg. within centroid distance_cluster_0: 0,815
Avg. within centroid distance_cluster_1: 0,973
Deviasi Bouldin: 4,399

Gambar 4. Hasil Perhitungan

4. KESIMPULAN

Berdasarkan hasil implementasi metode *K-Means Clustering* pada sentimen analisis produk SKINTIFIC maka penulis dapat menarik kesimpulan sebagai berikut:

1. Dengan menggunakan *software* RapidMiner untuk menganalisis data dengan cara crawling serta menerapkan metode *K-Means Clustering* pada data crawling untuk analisis sentimen produk penjualan shopee pada pengguna twitter menggunakan metode K-means.
2. Dari 159 data tweet yang digunakan, *cluster* yang memiliki jumlah *cluster* positif (C1) dengan persentase sebesar 79,5%. Sedangkan jumlah *cluster* negatif (C2) memiliki persentase sebesar 19,7%.

REFERENSI

Referensi utama adalah jurnal dan prosiding internasional. Semua referensi harus ke sumber yang paling relevan, terkini dan referensi minimum adalah 15. Referensi ditulis dalam gaya IEEE. Harap gunakan format yang konsisten untuk referensi - lihat contoh di bawah ini (11 pt):

- [1] Desta Yolanda, Mohammad Hafiz Hersyah, & Eno Marozi. (2021). Implementasi Metode Unsupervised Learning Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1099–1105. <https://doi.org/10.29207/resti.v5i6.3552>
- [2] Dewi, A. O. P. (2020). Kecerdasan Buatan sebagai Konsep Baru pada Perpustakaan. *Anuva: Jurnal Kajian Budaya, Perpustakaan, Dan Informasi*, 4(4), 453–460. <https://doi.org/10.14710/anuva.4.4.453-460>
- [3] Faesal, A., Muslim, A., Ruger, A. H., & Kusriani, K. (2020). Sentimen Analisis Terhadap Komentar Konsumen Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 207–213. <https://doi.org/10.30812/matrik.v19i2.640>
- [4] Farisi, M. H. Al. (2019). Analisis Sentimen Komentar Masyarakat Terhadap Kebijakan Pemerintah Tentang Sistem Zonasi Sekolah Menggunakan Algoritma K-Means Dan Algoritma Levensthein Distance. In *Institutional Repository UIN Syarif Hidayatullah Jakarta*.
- [5] Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66
- [6] Hadi, S. W., Julianto, M. F., Rahmatullah, S., & Gata, W. (2020). Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means. *Bianglala Informatika*, 8(2), 86–90. <https://doi.org/10.31294/bi.v8i2.8191>
- [7] Indriani, D. (2019). *Yayasan Lembaga Pendidikan Islam Daerah Riau Universitas Islam Riau Analisis Sentimen Pada Tweet Dengan Tagar # Kpujangancurang Menggunakan Metode Naïve Bayes*. [8] Nainggolan, Y. H., Manurung, S., & Rumapea, Y. (2022). Sumatera Utara 2) Program Studi Teknik Informatika. *Journal of Information and Technology Unimor*, 8, 49–52. <https://doi.org/10.32938/jitu.v2i2.2994>
- [8] Wahyuni, W. A., & Saepudin, S. (2021). Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci. *Seminar Nasional Sistem ...*, 306313. <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/35/31>
- [9] Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- [10] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- [11] Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film MenggunaRahutomo, F., Saputra, P. Y. and Fidyawan, M. A. (2018) ‘Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine’, *Jurnal Informatika Polinema. Jurnal Informatika Polinema*, 4(2), 93.
- [12] Malik, A. Z., Utami, E., & Raharjo, S. (2019). Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 dengan Metode K-NN. *Jurnal INFORMA Politeknik Indonesia Surakarta*, 5(2), 1–7.
- [13] Mala Olhang, M. M., Achmadi, S., & Wibisono, F. . A. (2020). Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (Nbc). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 4(2), 214–221. <https://doi.org/10.36040/jati.v4i2.2695>
- [14] Larasati, M. A. Z., Winarsih, N. A. S., Rohman, M. S., & Saraswati, G. W. (2022). Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter. *Progresif: Jurnal Ilmiah Komputer*, 18(2), 201. <https://doi.org/10.35889/progresif.v18i2.877>

-
- [15] Junior, A. R., Handayani, H. H., Fitri, A., & Masruriyah, N. (2022). *Analisis Sentimen Menggunakan Algoritma K-Means untuk Mengetahui Kalimat Positif maupun Negatif pada Buletin APTIKOM. III(1)*, 113.