



# Sentiment Analysis on User Reviews Emina Cheeklit Blusher Pressed Using the Naive Bayes Method

## Analisis Sentimen Pada Review Pengguna Emina Cheeklit Pressed Blush Dengan Menggunakan Metode Naive Bayes

*Silvyana Nur Haliza, Jamila Husna Nuraini\*, Rifqoh Andini, Cantika Indah Faradasya, Qory Andani Prasasti*

*Telkom University, Manajemen Bisnis Telekomunikasi & Informatika, Fakultas Ekonomi dan Bisnis*

Sociolla is a platform that is used to channel opinions and discuss various issues related to beauty. One of them is discussing beauty products. One of the products discussed is Emina Cheeklit Pressed Blush. Sentiment analysis is a way of comparing the opinions of many people. This study aims to determine the sentiment that occurs in every comment about Emina's Cheeklit Pressed Blush on the Sociolla platform. The sentiments used will be divided into 2 classes, positive and negative and using the Naive Bayes Classifier as the algorithm. The results obtained an accuracy of 77.69%.

### OPEN ACCESS

ISSN 2541-5107 (online)

**Edited by:**

*Akbar Wiguna*

**Reviewed by:**

*Bian Dwi Pamungkas*

**\*Correspondence:**

*Jamila Husna Nuraini  
milanuraini@student.telkomuniversit  
y.ac.id*

**Received:** 02-12-2020

**Accepted:** 17-12-2020

**Published:** 29-12-2019

**Citation:**

*Haliza SN, Nuraini JH, Andini R,  
Faradasya CI and Prasasti QA  
(2019) Sentiment Analysis on User  
Reviews Emina Cheeklit Blusher  
Pressed Using the Naive Bayes  
Method.*

*JICTE (Journal of Information and  
Computer Technology Education).  
3:2.*

*doi: 10.21070/jjcte.v3i2.187*

**Keywords:** Sentiment Analysis, Naive Bayes Classifier

Sociolla merupakan salah satu platform yang digunakan untuk menyalurkan opini dan membahas berbagai masalah yang berkaitan dengan kecantikan. Salah satunya adalah membahas mengenai produk kecantikan. Salah satu produk yang dibahas adalah Emina Cheeklit Pressed Blush. Analisis sentimen merupakan suatu cara untuk membandingkan pendapat banyak orang. Penelitian ini bertujuan untuk menentukan sentimen yang terjadi pada setiap komentar mengenai Emina Cheeklit Pressed Blush pada platform Sociolla. Sentimen yang digunakan akan dibagi menjadi 2 kelas, yaitu positif dan negatif serta menggunakan Naive Bayes Classifier sebagai algoritmanya. Hasil penelitian diperoleh akurasi 77,69%.

**Keywords:** Analisis Sentimen, Naive Bayes Classifier

## PENDAHULUAN

Industri kecantikan berkembang begitu pesat dalam beberapa tahun ke belakang. Hal tersebut dapat dilihat dari berbagai macam tren kecantikan yang turut memengaruhi pola konsumen di Indonesia. Berdasarkan survei dari ZAP Beauty Index 2018 terhadap 17.889 perempuan di Indonesia secara online menemukan beberapa fakta soal industri kecantikan, yaitu diantaranya perempuan di Indonesia sudah mulai mengenal *make up* ketika usia mereka kurang dari 18 tahun (13-15 tahun) atau sekitar 41,9 persen dari hasil survei yang telah dilakukan. Dari hal tersebut dapat dilihat bahwa industri kecantikan mulai dikenal oleh anak-anak usia muda atau remaja, terutama pada Generasi Y dan Z [1]. Berdasarkan survei yang sudah dilakukan, terdapat *top 5 most talked about brands on twitter*. Didapatkan lima nama brand yang paling sering dibicarakan netizen dapat dilihat pada Gambar 1

Merek Kosmetik	Posts
Wardah	500
Sariayu	399
Purbasari	278
Emina	238
Mustika Ratu	75

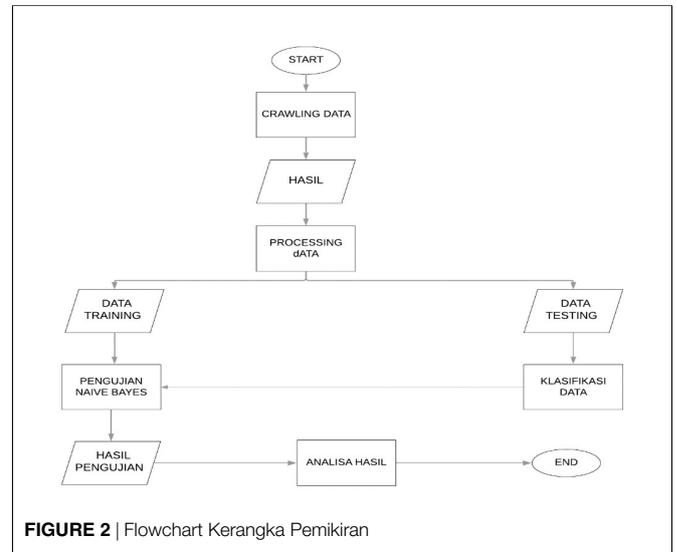
**FIGURE 1** | Top 5 most talked about brands on twitter

Salah satu *brand* produk kecantikan dari Indonesia yang memiliki target pasar perempuan berusia remaja adalah Emina. Meski Emina masih tergolong produk baru dalam industri kosmetik lokal, tetapi emina sudah berhasil menduduki peringkat 5 besar. Hal tersebut menunjukkan bahwa para remaja memiliki ketertarikan pada produk Emina. Dari berbagai produk yang dimiliki oleh Emina, produk *Emina Cheeklit Pressed Blush* merupakan salah satu produk *blush on* yang paling digemari terutama oleh kaum remaja perempuan, dilihat berdasarkan *review* dari yang sudah pernah menjadi konsumen *Emina Cheeklit Pressed Blush*. Hal ini dapat dimanfaatkan untuk mengumpulkan informasi dan data dalam melakukan analisis sentimen. Data yang akan dikumpulkan pada penelitian ini merupakan *review* dari pengguna *Emina Cheeklit Pressed Blush* yang diambil dari web *sociolla.com* dengan menggunakan *tools* *parsehub* dan diolah menggunakan metode *Naive Bayes Classifier*.

## METODE PENELITIAN

### Kerangka Pemikiran

Pada penelitian ini akan diberikan gambaran langkah-langkah dalam mendeteksi kelas sentimen menggunakan metode *Naive Bayes*. Berikut gambaran dari proses mendeteksi kelas sentimen dengan metode *Naive Bayes* yang akan digunakan dapat dilihat pada Gambar 2



Pada Gambar 2 dapat dilihat bahwa *Crawling data* dari *Sociolla review Emina Cheeklit Pressed Blush* dengan menggunakan *tools* *parsehub*. Setelah mendapatkan hasil *crawling data*, dilakukan *pre-processing* secara manual untuk membersihkan data mentah menjadi data yang terstruktur. Lalu menentukan data *training* dan data *testing* yang sudah di *pre-processing* secara manual. Pada data *training* dilakukan *klasifikasi data* untuk menentukan sentimen Pang and Lee (2008). Langkah selanjutnya dilakukan *pengujian Naive Bayes Classifier* terhadap data *training* yang sudah *diklasifikasi* dan data *testing* Prasetyo (2012) Pratiwi and Nugroho (2016). Setelah mendapatkan hasil dari *pengujian*, hasil dari *pengujian Naive Bayes Classifier* di Analisa Nugroho and Setyawan (2014).

### Crawling Data

*Crawling Data* adalah suatu metode untuk mengumpulkan atau mengambil data dari suatu blog, sosial media atau situs lainnya Saleh (2015). Dalam hal ini *Sociolla* menjadi website untuk pengambilan dataset. *Crawling* dilakukan dengan menggunakan *tools* *parsehub* terkait tentang *review* tingkat kepuasan konsumen terhadap produk *Emina Cheeklit Pressed blush*. *Review* yang diambil menggunakan Bahasa Indonesia, total data yang akan diambil sebanyak 1000 *review*. Contoh *review* hasil *crawling* ditampilkan pada Tabel 1 yang diambil secara acak dari data asli.

**TABLE 1** | Sample Data

No	Review
1.	Aku beli blush emina ini yg shade cotton candy. Bagus banget warna nya, pink2 natural gitu. Pigmentasi nya jg bagus banget, apalagi harganya jg terjangkau bgt.
2.	Buat hari2 ok seh..imut n gampang dibawa2. Tp klo tuk acara2 pesta atau special kyaknya kurang mantul deh..heheh. Baru punya 1 shade. Mgkn lain x bisa beli shade yg lain.
3.	Bentuknya kecil jadi ga banyak makan tempat, warnanya cantik gampang di aplikasikan jg aku pake yg marshmallow lady cocok buat yg punya kulit kuning langsung, ketahanannya jg oke recommend lah.

**TABLE 3** | Contoh Sample Data sebelum diberikan label

No	Review
1.	Aku beli blush emina ini yg shade cotton candy. Bagus banget warna nya, pink2 natural gitu. Pigmentasi nya jg bagus banget, apalagi harganya jg terjangkau bgt.
2.	Buat hari2 ok seh..imut n gampang dibawa2. Tp klo tuk acara2 pesta atau special kyaknya kurang mantul deh..heheh. Baru punya 1 shade. Mgkn lain x bisa beli shade yg lain.
3.	Bentuknya kecil jadi ga banyak makan tempat, warnanya cantik gampang di aplikasikan jg aku pake yg marshmallow lady cocok buat yg punya kulit kuning langsung, ketahanannya jg oke recommend lah.

## Preprocessing Data

Sebelum dilakukan pengolahan, data yang disimpan perlu di processing terlebih dahulu untuk “dibersihkan” [Widiastuti et al. \(2014\)](#) . Pre-processing adalah tahapan dimana data atau informasi diatur dari data tekstual semi terstruktur dan tidak terstruktur menjadi sebuah format yang lebih terstruktur yang nantinya akan mempermudah pemakai. Kemudian data tersebut akan terbagi menjadi dua yaitu data training dan data testing [Indraswari and Kurniawan \(2018\)](#) .

**TABLE 2** | Preprocessing Data

No	Review
1.	Aku beli blush emina ini yg shade cotton candy. Bagus banget warna nya, pink2 natural gitu. Pigmentasi nya jg bagus banget, apalagi harganya jg terjangkau bgt.
2.	Buat hari2 ok seh..imut n gampang dibawa2. Tp klo tuk acara2 pesta atau special kyaknya kurang mantul deh..heheh. Baru punya 1 shade. Mgkn lain x bisa beli shade yg lain.
3.	Bentuknya kecil jadi ga banyak makan tempat, warnanya cantik gampang di aplikasikan jg aku pake yg marshmallow lady cocok buat yg punya kulit kuning langsung, ketahanannya jg oke recommend lah.

Berikut ini merupakan tahapan yang dilakukan pada pre-processing tweet pada Tabel 2 :

1. Tokenizing merupakan proses untuk memisahkan kalimat pada data menjadi beberapa kata.
2. Cleansing adalah proses pembersihan data teks yaitu dengan menghilangkan data yang tidak konsisten atau tidak relevan seperti emoticon, hashtag (#), username, url.
3. Normalization digunakan untuk menormalisasikan kalimat gaul menjadi kalimat yang sesuai dengan KBBI.
4. Case folding yaitu merubah bentuk kata yang memiliki imbuhan menjadi bentuk dasarnya.

## Labelling

Pada proses *labelling* data yang sudah dilakukan proses pre-processing diberikan label secara manual [Marlina \(2017\)](#) . Label ini digunakan sebagai klasifikasi kelas pada setiap data. Kelas yang digunakan yaitu kelas positif dan kelas negatif. Tabel 3 dan Tabel 4 menunjukkan contoh data yang diambil secara acak pada data *Emina Cheeklit Pressed blush* sebelum dan sesudah diberikan label.

**TABLE 4** | Contoh Sample Datasudah diberikan label

No	Review	Sentimen
1.	Aku beli blush emina ini yg shade cotton candy. Bagus banget warna nya, pink2 natural gitu. Pigmentasi nya jg bagus banget, apalagi harganya jg terjangkau bgt.	Positif
2.	Buat hari2 ok seh..imut n gampang dibawa2. Tp klo tuk acara2 pesta atau special kyaknya kurang mantul deh..heheh. Baru punya 1 shade. Mgkn lain x bisa beli shade yg lain.	Positif
3.	Bentuknya kecil jadi ga banyak makan tempat, warnanya cantik gampang di aplikasikan jg aku pake yg marshmallow lady cocok buat yg punya kulit kuning langsung, ketahanannya jg oke recommend lah.	Positif

## HASIL DAN PEMBAHASAN

Dari pengujian yang sudah dilakukan dalam mencari probabilitas, confusion matrix, akurasi, precision, recall, dan f-measure didapatkan nilai seperti berikut:

### Hasil Pengujian Probabilitas

**TABLE 5** | Probabilits Data Training

Kelas	Banyak Data	Probabilitas Kelas
Positif	588	0,841
Negatif	111	0,158
Jumlah	699	

**TABLE 6** | Probabilits Data Testing

Kelas	Banyak Data	Probabilitas Kelas
Positif	154	0,527
Negatif	138	0,472
Jumlah	292	

Dari Tabel 5 dan Tabel 6 diatas, dapat disimpulkan bahwa *ReviewEmina Cheeklit Pressed Blush* pada data *training* mempunyai probabilitas positif sebesar 0,841 dan negatif sebesar 0,158, sedangkan pada data *testing* mempunyai probabilitas positif sebesar 0,527 dan negatif sebesar 0,472.



- gunakan algoritma c4.5. komunitas. vol. VI. 84–91.
- Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends\* in Information Retrieval* 2, 1–135. doi: 10.1561/1500000011.
- Prasetyo, E. (2012).
- Pratiwi, R. and Nugroho, Y. S. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro* 8, 60–63.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal* 2, 207–217.
- Widiastuti, N. A., Santosa, S., and Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Particle Swarm Optimization Untuk Deteksi Penyakit Jantung. *Jurnal Pseudocode* 1, 11–14.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Haliza, Nuraini, Andini, Faradasya and Prasasti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.