

Sentiment Analysis on Emina Cheeklit Pressed Blush User Review Using the Naive Bayes Method

Analisis Sentimen pada Review Pengguna Emina Cheeklit Pressed Blush dengan Menggunakan Metode Naive Bayes

<i>Silvyana Nur Haliza</i>	5S1 MBTI, Fakultas Ekonomi dan Bisnis, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu, Bandung , Jawa Barat, Indonesia 40257
<i>Jamila Husna Nuraini</i>	S1 MBTI, Fakultas Ekonomi dan Bisnis, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu, Bandung , Jawa Barat, Indonesia 40257
<i>Rifqoh Andini</i>	S1 MBTI, Fakultas Ekonomi dan Bisnis, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu, Bandung , Jawa Barat, Indonesia 40257
<i>Cantika Indah Faradasya</i>	S1 MBTI, Fakultas Ekonomi dan Bisnis, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu, Bandung , Jawa Barat, Indonesia 40257
<i>Qory Andani Prasasti</i>	S1 MBTI, Fakultas Ekonomi dan Bisnis, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu, Bandung , Jawa Barat, Indonesia 40257

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

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 survey yang sudah dilakukan, terdapat *top 5 most talked about brands on twitter*. Didapatkan lima nama brand yang paling sering dibicarakan netizen dapat dilihat pada [Figure 1](#)

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 [Figure 2](#)

Figure 2. *Flowchart Kerangka Pemikiran*

Pada [Figure 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 sentimen¹. Langkah selanjutnya dilakukan pengujian *Naive Bayes Classifier* terhadap data *training* yang sudah diklasifikasi dan data *testing*^{2 3}. Setelah mendapatkan hasil dari pengujian, hasil dari pengujian *Naive Bayes Classifier* di Analisa⁴.

Crawling Data

Crawling Data adalah suatu metode untuk mengumpulkan atau mengambil data dari suatu blog, sosial media atau situs lainnya⁵. 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 [Table 1](#) yang diambil secara acak dari data asli.

Table 1. *Sample Data*

Preprocessing Data

Sebelum dilakukan pengolahan, data yang disimpan perlu di *processing* terlebih dahulu untuk

“dibersihkan”⁶ . 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 testing⁷ .

Table 2. *Preprocessing Data*

Berikut ini merupakan tahapan yang dilakukan pada preprocessing tweet pada **Table 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 preprocessing diberikan label secara manual⁸ . Label ini digunakan sebagai klasifikasi kelas pada setiap data. Kelas yang digunakan yaitu kelas positif dan kelas negatif. **Table 3** dan **Table 4** menunjukkan contoh data yang diambil secara acak pada data *Emina Cheeklit Pressed blush* sebelum dan sesudah diberikan label.

Table 3. *Contoh Sample Data sebelum diberikan label*

--	--	--

Table 4. Contoh Sample Datas sudah diberikan label

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

Table 6. Probabilits Data Testing

Dari **Table 5** dan **Table 6** diatas, dapat disimpulkan bahwa *Review Emina 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.

Hasil Pengujian Performansi

Table 7. Performansi Data Training

Dari **Table 7** dapat dihitung nilai akurasi, *recall*, *precision* dan *f-measure* berdasarkan sentimen

positif dan hasilnya seperti terlihat pada **Table 8** :

Table 8. Nilai Akurasi, Recall, Precision, F-Measure

Dari **Table 7** , merupakan nilai akurasi, recall, precision, f-measure berdasarkan sentiment positifnya.

Dari **Table 8** dapat disimpulkan bahwa *Review Emina 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. Dari *table* diatas, merupakan nilai akurasi, recall, precision, f-measure berdasarkan sentiment positifnya.

Figure 3. Presentasi Proporsi Data Emina Cheeklit Pressed Blush

Figure 4. Hasil Pengolahan WordCloud

Dapat dilihat pada **Figure 3** , presentase sentimen positif sebesar 74%, dan sentimen negatif sebesar 26%. Hal tersebut dapat menjelaskan bahwa *Emina Cheeklit Pressed Blush* merupakan produk yang mendapatkan respon positif dari customer, hal ini didukung oleh berbagai faktor yaitu seperti warna *blush* yang bervariasi, pigmentasi produk sangat bagus, cocok digunakan untuk kulit perempuan Indonesia, hingga harganya yang dapat dikategorikan murah. Selain itu, kekurangan dari produk *Emina Cheeklit Pressed Blush* menurut **Figure 4** yaitu pada packaging yang ringkih dan mudah pecah.

Kesimpulan

Dari analisis percobaan yang dilakukan terkait sentimen terhadap *Emina Cheeklit Pressed Blush* dari Sociolla dengan metode Naive Bayes. Maka dapat disimpulkan bahwa pada penelitian ini didapatkan akurasi sebesar 77,69% yang diperoleh dari pengujian data dengan menggunakan algoritma Naive Bayes dengan tools RapidMiner. proses klasifikasi akan semakin akurat jika data training yang digunakan semakin banyak. proses pelabelan data menjadi dasar untuk membangun model yang lebih baik. *Emina Cheeklit Pressed* merupakan produk yang mendapatkan respon positif dari customer, hal ini didukung oleh berbagai faktor yaitu seperti warna *blush* yang bervariasi, pigmentasi produk sangat bagus, cocok digunakan untuk kulit perempuan Indonesia, hingga harganya yang dapat dikategorikan murah. Selain itu, kekurangan dari produk *Emina Cheeklit Pressed Blush* menurut data diatas yaitu pada packaging yang ringkih dan mudah pecah.

Dari hasil pengolahan data tersebut, perusahaan juga dapat memperbaiki serta meningkatkan kualitas packaging pada produk Emina Cheeklit Pressed Blush agar lebih kuat, tidak ringkih serta tidak mudah pecah.

References