

Analisis Sentimen terhadap Produk Skin Game di Forum *Review Female Daily* Menggunakan Metode *Multinomial Naïve Bayes* dan TF-IDF

Quina Alifa Nanda Prakoso¹, Anita Muliawati², Ika Nurlaili Isnainiyah³

^{1,2,3} Informatika, Fakultas Ilmu Komputer

^{1,2,3} Universitas Pembangunan Nasional Veteran Jakarta

^{1,2,3} Jl. RS Fatmawati No. 1, Pondok Labu, Jakarta Selatan DKI Jakarta 12450

quinaalifa@upnvj.ac.id¹, anitamuliawati2017prodi@gmail.com², nurlailika@upnvj.ac.id³

Abstrak. *Female Daily* merupakan salah satu *platform online* yang berfokus di bidang kecantikan dan memiliki forum *review* dengan lebih dari 50.000 produk kosmetik yang dapat *dirีวิว* oleh *membrny*. Salah satu *brand* yang berada di *platform* ini adalah *Skin Game*, *brand* kosmetik lokal yang didirikan pada tahun 2019. Dalam proses pemilihan *skincare*, seringkali calon pembeli membutuhkan opini orang, oleh karena itu *Female Daily* hadir memenuhi kebutuhan tersebut. Namun dengan banyaknya *review* yang tersedia, pengguna membutuhkan waktu untuk membaca beberapa *review* terlebih dahulu. Cara lain yang dapat dilakukan untuk mempermudah hal tersebut adalah klasifikasi polaritas sentimen pada *review* produk. Dalam penelitian ini, 450 data *review* produk *Acne Warrior* dari *brand* *Skin Game* diambil melalui forum *review* *Female Daily* kemudian diklasifikasi menggunakan metode *Multinomial Naïve Bayes* dan TF-IDF untuk mengetahui sentimennya. Hasil evaluasi model pada penelitian ini mendapatkan *accuracy* sebesar 86,6%, *precision* 89,1%, *recall* 94,2%, dan *specificity* 60%. Hasil penelitian ini berupa polaritas sentimen terhadap produk *Skin Game*. Berdasarkan hasil pelabelan, dari 450 *review*, 342 *review* (76%) memiliki sentimen positif yang artinya cocok serta merasakan manfaat dari produk tersebut dan 108 *review* (24%) memiliki sentimen negatif yang artinya tidak cocok atau produk *Acne Warrior Skin Game* tidak memenuhi ekspektasi pengguna.

Kata Kunci: *Multinomial Naïve Bayes*, TF-IDF, Analisis Sentimen, *Skin Game*, *Female Daily*.

1 Pendahuluan

Kosmetik merupakan kebutuhan manusia, baik wanita maupun pria. Menurut buku A-Z tentang Kosmetik pada dasarnya kosmetik merupakan campuran bahan yang dioleskan ke bagian luar tubuh seperti lapisan terluar kulit, kuku, rambut, bibir, gigi, dan lainnya untuk tujuan menambah pesona, perlindungan, perbaikan, agar penampilannya lebih cantik dari sebelumnya [1]. Kosmetik merupakan salah satu sektor industri yang justru mengalami pertumbuhan positif selama pandemi Covid-19, terutama di tahun 2021 ini. Berdasarkan data bulan Januari 2021 dari *dashboard* *Compas.co.id*, kategori perawatan kecantikan mengungguli transaksi penjualan di *e-commerce* sebesar 46.8% [2]. 50% dari *brand* kecantikan wajah yang masuk ke peringkat 10 besar berdasarkan penjualan di bulan Januari ini merupakan *brand* lokal, data ini menunjukkan bahwa *brand* lokal dapat bersaing dengan *brand* global.

Skin Game adalah salah satu *brand* kosmetik lokal yang menawarkan produk-produk untuk mengatasi jerawat serta komedo [3]. *Brand* ini meluncurkan produk pertamanya di tahun 2019 dan hingga saat ini, *Skin Game* sudah meluncurkan 10 produk *skincare* dan 3 produk *make-up*. Produk tersebut terdiri dari obat total jerawat, masker untuk komedo, masker *clay*, *essence*, *serum*, *cleanser* dan *toner*. Produk pertama *Skin Game* yaitu *Acne Warrior* (obat total jerawat) sudah terjual lebih dari 40.000 pcs di bulan Juni 2021.

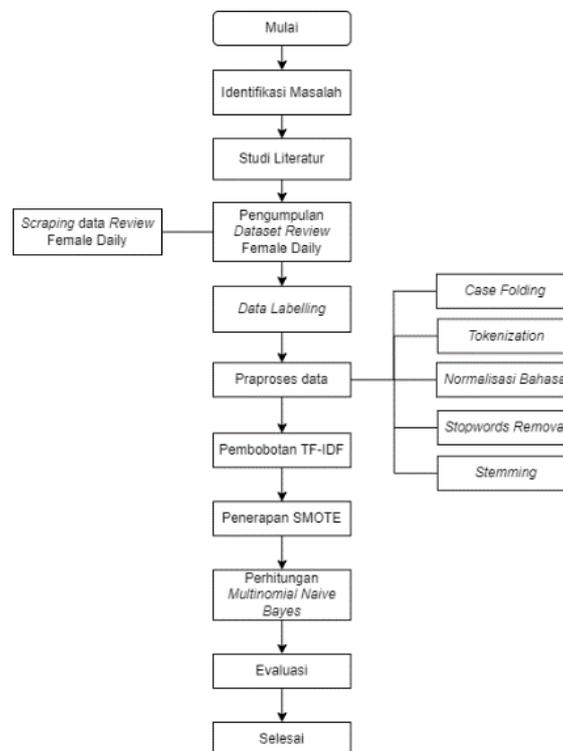
Dalam studi berjudul *Connected Beauty Consumer Report*, 26% konsumen usia 18-35 tahun membeli produk selepas melihat *review* dan sebanyak 40% memakai pencarian di internet untuk melihat *review* produk [4]. Saat ini sudah banyak media sosial yang dapat menyalurkan opini mulai dari Twitter, Instagram, Facebook, Blog, Youtube dan platform khusus *review* produk kosmetik, salah satunya *Female Daily*. *Female Daily* merupakan destinasi kecantikan nomor satu di Indonesia yang memuat artikel serta *review* produk kecantikan dari 865.000 member di seluruh Indonesia. Di setiap subkategori produk, kita dapat mengurutkan produk tersebut berdasarkan beberapa kategori, salah satunya kepopuleran produk. Produk populer di *reviews.femaledaily.com* merupakan produk yang memiliki *review* paling banyak, tetapi tidak berarti produk tersebut mempunyai sentimen positif. Apabila sentimen produk tersebut negatif maka kemungkinan besar produk tersebut tidak direkomendasikan atau tidak bagus untuk digunakan walaupun mempunyai *review* yang banyak. Oleh karena itu, perlu dilakukan klasifikasi sentimen pada *review* produk untuk memahami produk tersebut bagus serta direkomendasikan atau tidak.

Klasifikasi sentimen memerlukan algoritma seperti *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Maximum Entropy* (ME), dan *K-Nearest Neighbor* (KNN). Algoritma ini digunakan dalam beberapa penelitian seperti analisis sentimen *review* produk kosmetik [5], *movie reviews* [6], komentar peserta diklat [7], dan *hotel reviews* [8]. *Naïve Bayes* merupakan model yang dikenal simpel dan efektif [5] serta memiliki asumsi yang kuat terhadap independensi dari setiap kondisi, terutama ketika data latih yang digunakan sedikit. Pada penelitian yang dilakukan oleh [6] diuji 5 *classifier* untuk analisis sentimen *movie review* diantaranya *Bernoulli Naïve Bayes* (BNB), *Decision Tree* (DE), *Support Vector Machine* (SVM), *Maximum Entropy* (ME), serta *Multinomial Naïve Bayes* (MNB) dan hasil akurasi tertinggi sebesar 88.5% dihasilkan oleh *Multinomial Naïve Bayes classifier*. Oleh karena itu, penelitian ini menggunakan *Multinomial Naïve Bayes classifier*.

Fitur pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*) dapat melihat seberapa relevan suatu kata pada dokumen [9]. Pada penelitian analisis sentimen berbasis aspek oleh [10] dibandingkan dua model yaitu POS *tagging* dan TF-IDF pada 2 *classifier*. Hasil penelitian ini menunjukkan performansi TF-IDF lebih baik dibanding POS *tagging* pada kedua *classifier*, *Naïve Bayes* dan *Support Vector Machine*. Pada *classifier Naïve Bayes* didapatkan akurasi sebesar 75.71%. Studi lainnya yaitu *Sentiment Analysis of the Enforcement of PSBB Part II in Jakarta* di media sosial Twitter oleh [11] membandingkan 2 *classifier* yaitu *K-Nearest Neighbor* dan *Multinomial Naïve Bayes* dengan pembobotan TF-IDF. Dari hasil evaluasi didapatkan akurasi sebesar 85% untuk *Multinomial Naïve Bayes classifier* dengan pembobotan TF-IDF. Berdasarkan beberapa penelitian yang sudah dipaparkan, peneliti mengusulkan pengklasifikasian sentimen *review* produk brand Skin Game menggunakan metode *Multinomial Naïve Bayes* dan TF-IDF dengan keinginan menghasilkan akurasi yang lebih baik.

2 Metodologi Penelitian

Tahapan yang dilakukan dalam penelitian ini tersaji dalam Gambar 1 di bawah ini:



Gambar. 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Sebelum melakukan sebuah penelitian, perlu dilakukan identifikasi masalah, yaitu proses upaya menjelaskan masalah yang akan diteliti dan dicari solusinya. Secara singkat, identifikasi adalah mendefinisikan masalah dalam penelitian.

2.2 Studi Literatur

Studi literatur merupakan tahapan mencari referensi teori yang relevan dan berkaitan dengan permasalahan yang ditemukan di tahapan sebelumnya. Dalam penelitian ini riset yang dilakukan mengacu pada jurnal yang berkaitan dengan *Text Mining*, *Algoritma Naïve Bayes*, dan TF-IDF untuk dijadikan referensi penelitian.

2.3 Pengumpulan Dataset Review Female Daily

Data dikumpulkan dengan menggunakan *Web Scraping* kolom Review di website reviews.femaledaily.com dan diseleksi secara manual untuk menghilangkan *review* yang menggunakan bahasa Inggris secara keseluruhan. Data yang diambil adalah empat produk *skincare* dari brand Skin Game paling populer versi reviews.femaledaily.com, dengan menggunakan fitur '*sort by: Popular*'. Fitur ini menampilkan produk sesuai dengan urutan jumlah *review*, produk yang paling banyak *direview* akan menempati posisi teratas, diikuti oleh produk dengan jumlah *review* lebih sedikit dari produk teratas. Data produk yang diambil adalah produk dengan jumlah *review* terbanyak, yaitu Acne Warrior (765 reviews).

2.4 Data Labelling

Data labelling dilakukan oleh tiga penilai secara manual dan dikategorikan menjadi dua label yaitu positif dan negatif. Pelabelan melibatkan tiga penilai agar menemukan kesepakatan apabila terdapat penilaian yang berbeda.

Tabel 1. Contoh Pelabelan Data *Review*

Data <i>Review</i>	Penilai 1	Penilai 2	Penilai 3	Label Akhir
BEST ACNE SPOT TREATMENTTTT gatau harus ngomong apa lagi tapi aku udah purchase sampe 3 kali karena SEBAGUS ITU□□□ biasanya acne spot lainnya kalo dipake pasti jadi berbekas hitam tp skin game acne warrior ngga!! dari segi harga agak mahal tp worth the price soalnya beneran ampuh□	Positif	Positif	Positif	Positif
GILA. Ini the best product banget sih untuk semua jenis jerawat di aku! Produk ini tuh nggak ngempesin jerawat melainkan memantengkan jerawat. Itu justru yang aku suka T_T Di aku ini nggak bikin jerawat berbekas hitam. Untuk performa, jerawat jenis nodul kempes dalam waktu 3-4 hari. Aku pakai ini selama di rumah aja di pagi dan malam.	Positif	Positif	Negatif	Positif
Sebenarnya ini ngaruh di jerawat meradangku, cuma ya emang harus telaten pakai tiap malam dan tipis2 aja yaa. Bisa memicu iritasi sekitaran soalnya. Selain itu, kalau dipakai tebal dikit bisa nempel2 ke bantal kalo dipake tidur pagi2 dah kegeser2. Di aku, dia gabisa kasih efek yg instan sih. Misal jerawat meradang besar tu butuh bgt beberapa hari sampai dia mau kempis. Kalau yg kecil2 bisa lumayan cepet ilangnya .	Negatif	Positif	Negatif	Negatif

Selain itu, dilakukan juga uji statistik kappa untuk melihat reliabilitas kesepakatan para penilai. Kappa Fleiss dihitung menggunakan persamaan (1) dan hasil nilai dengan rentang 0,00 – 1,00.

$$kappa\ Fleiss = \left[\frac{P(a) - P(e)}{1 - P(e)} \right] \tag{1}$$

Keterangan:

P(a) : nilai kesepakatan seluruh penilai

P(e) : peluang kesepakatan sejumlah penilai yang penilaiannya berbeda

Dimana P(e) dapat dihitung dengan persamaan (2).

$$P_e = P(positif)^2 + P(negatif)^2 \tag{2}$$

Nilai kappa Fleiss yang sudah dihitung kemudian diukur berdasarkan interpretasi nilai kappa, dimana nilai $< 0,00$ diinterpretasikan sebagai *poor*, $0,00 - 0,20$ *slight*, $0,21 - 0,40$ *fair*, $0,41 - 0,60$ *moderate*, $0,61 - 0,80$ *substantial*, dan $> 0,81$ *almost perfect*. Berdasarkan interpretasi tersebut, dapat disimpulkan bahwa semakin besar nilai kappa, semakin baik hasilnya karena para penilai memiliki kesepakatan yang tidak terlalu berbeda.

2.5 Praproses Data

Praproses data ialah tahap mempersiapkan data mentah yang belum terstruktur menjadi data yang siap digunakan untuk tahap berikutnya. Pada riset ini, proses yang hendak penulis lakukan adalah *Case Folding*, *Tokenization*, *Normalisasi Bahasa*, *Stopwords Removal* dan *Stemming*.

- Case Folding**, tahapan praproses data di mana seluruh huruf dalam *review* produk akan diganti menjadi huruf kecil. Karakter selain huruf kecil seperti tanda baca juga akan dihapuskan.
- Tokenization**, setelah diubah seluruh huruf menjadi huruf kecil dan membersihkan data, tahapan selanjutnya adalah membagi sebuah teks menjadi token-token.
- Normalisasi Bahasa**, teks yang telah terbagi menjadi token-token kemudian diubah dari kata tidak baku menjadi baku, dan disesuaikan dengan KBBI.
- Stopwords Removal**, data *review* yang telah diubah Kembali menjadi baku kemudian dilanjutkan dengan proses *stopwords removal* yaitu, meniadakan kata yang sering muncul namun tidak berpengaruh pada performa klasifikasi.
- Stemming**, Terakhir untuk pra proses data dilakukan *stemming* yaitu, mentransformasi kata dalam *review* produk menjadi *root* kata menggunakan *library* Sastrawi dengan melenyapkan imbuhan.

2.6 Pembobotan TF-IDF

Tahap berikutnya ialah pembobotan kata menggunakan penghitungan TF-IDF (*Term Frequency – Inverse Document Frequency*). TF-IDF melalui tahapan yang dirumuskan sebagai berikut [11].

- Hitung bobot *Inverse Document Frequency* (IDF)

$$idf_t = \log \frac{N}{df_t} \quad (3)$$

- Hitung bobot TF-IDF

$$W_{t,d} = Wt_{t,d} \times idf_t \quad (4)$$

Keterangan:

idf_t = Nilai IDF kata- t

N = Jumlah dokumen.

df_t = Jumlah dokumen yang mengandung kata- t .

$W_{t,d}$ = Hasil pembobotan TF-IDF.

$Wt_{t,d}$ = Nilai kemunculan kata- t di dalam dokumen- d .

2.7 Penerapan SMOTE

Data yang diambil seringkali memiliki ketimpangan antar kelasnya. Untuk menyiasati data yang *imbalance* maka dilakukan penerapan SMOTE yaitu *Synthetic Minority Over-Sampling Technique*. Teknik SMOTE ini digunakan untuk menyeimbangkan data antara *review* positif dan negatif. Pertama, *dataset* yang sudah diolah akan dibagi menjadi data latih dan data uji terlebih dahulu kemudian dilakukan penerapan SMOTE untuk data latih [12]. Hasilnya berupa data latih awal yang menjadi minoritas ditambah dengan hasil *oversampling* dengan SMOTE.

2.8 Perhitungan Multinomial Naïve Bayes

Tahap selanjutnya adalah perhitungan *Multinomial Naïve Bayes*. Perhitungan ini dilakukan guna mengklasifikasi sentimen menjadi kelas positif dan negatif. Tahapan yang akan dilakukan dalam perhitungan sebagai berikut [11]

- Menentukan nilai *prior* kelas dengan persamaan (5).

$$P(c) = \frac{N_c}{N} \quad (5)$$

Keterangan:

N_C = Jumlah dokumen kelas c pada seluruh dokumen.
 N = Jumlah dokumen.

b. Menentukan probabilitas bersyarat tiap kata dengan persamaan (6).

$$P(t_n|c) = \frac{W_{ct} + 1}{(\sum W' \in V W'_{ct}) + B'} \quad (6)$$

Keterangan:

W_{ct} = Nilai pembobotan TF-IDF dari *term* t di kelas c
 $\sum W' \in V W'_{ct}$ = Jumlah semua *term* yang berada di kelas c
 B' = Jumlah *term* unik pada semua dokumen

c. Menghitung probabilitas suatu dokumen masuk ke suatu kelas dengan persamaan (7).

$$P(c|term\ dokumen\ d) = P(c) \times P(t_1|c) \times P(t_2|c) \times P(t_3|c) \times \dots \times P(t_n|c) \quad (7)$$

Keterangan:

$P(c)$: Probabilitas prior dari kelas c.

t_n : Kata dokumen d ke-n

$P(c|term\ dokumen\ d)$: Probabilitas suatu dokumen termasuk kelas c

$P(t_n|c)$: Probabilitas kata ke-n dengan diketahui kelas c

d. Menentukan kelas dokumen dengan memilih nilai probabilitas tertinggi.

2.9 Evaluasi

Tahap terakhir pada penelitian ini adalah evaluasi yang dilakukan dengan membangun *confusion matrix*. *Accuracy*, yaitu rasio prediksi benar (positif dan negatif) terhadap seluruh data dapat dihitung menggunakan persamaan (8). *Precision*, yaitu rasio prediksi benar positif terhadap seluruh hasil yang diprediksi positif dapat dihitung menggunakan persamaan (9). *Recall*, yaitu rasio prediksi benar positif terhadap seluruh data yang benar positif dapat dihitung menggunakan persamaan (10). *Specificity*, yaitu rasio prediksi benar negatif terhadap seluruh data negatif dapat dihitung menggunakan persamaan (11).

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

Keterangan:

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative

3 Hasil Pembahasan

Pengambilan data dilakukan menggunakan ekstensi *website* bernama Data Miner dan didapatkan 765 data pada tanggal 30 April 2022. Data hanya diambil satu kali dan dilakukan pengambilan semua data *review* yang tersedia untuk produk Acne Warrior dari brand Skin Game. Dari 765 data, diambil 450 data yang disaring secara manual. Data yang tidak dipilih merupakan *review* berbahasa Inggris secara mayoritas maupun keseluruhan, serta *review* yang hanya mendeskripsikan produk tanpa memberikan sentimen negatif ataupun positif. Setelah disaring, dilakukan pelabelan manual, data diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Penentuan ini dilakukan oleh 3 orang penilai dengan mempertimbangkan performa produk, aspek seperti harga, aroma, dan sebagainya tidak dijadikan acuan. Dari pelabelan data tersebut didapatkan 342 *review* positif dan 108 *review* negatif.

Data yang telah dilabeli kemudian masuk ke tahap pra proses data yang terdiri dari *case folding*, *tokenization*, normalisasi bahasa, *stopwords removal*, dan *stemming*. Hasil pra proses dapat dilihat pada tabel di bawah. Contoh hasil pra proses di bawah merupakan data latih yang akan digunakan.

Tabel 2. Hasil Pra Proses Data Latih

Review
['ampuh', 'jerawat', 'pendam', 'salah', 'acne', 'spot', 'bagus', 'jerawat', 'bukti', 'tinggal', 'bekas', 'hitam', 'wajah', 'jerawat', 'butuh', 'mantap', 'hasil']
['pakai', 'pengaruh', 'udah', 'lama', 'lama', 'enggak', 'tau', 'tidak', 'mempan', 'jerawat', 'senang', 'cari', 'akhir', 'cari', 'obat', 'totol', 'jerawat', 'enak', 'kering', 'kulit']

Selanjutnya masuk ke tahap pembobotan TF-IDF. Setiap kata pada tabel 2, dihitung frekuensinya. Tabel perhitungannya dapat dilihat pada tabel 3.

Tabel 3. Pembobotan Kata TF-IDF

Term	Dokumen		DF	IDF	TF-IDF	
	D1	D2			D1	D2
lapor	1	0	1	0,30103	0,30103	0
ampuh	3	1	2	0	0	0
jerawat	1	0	1	0,30103	0,30103	0
pendam	1	0	1	0,30103	0,30103	0
salah	1	0	1	0,30103	0,30103	0
acne	1	0	1	0,30103	0,30103	0
spot	1	0	1	0,30103	0,30103	0
bagus	1	0	1	0,30103	0,30103	0
bukti	1	0	1	0,30103	0,30103	0
bekas	1	0	1	0,30103	0,30103	0
hitam	1	0	1	0,30103	0,30103	0
wajah	1	0	1	0,30103	0,30103	0
butuh	1	0	1	0,30103	0,30103	0
mantap	1	1	1	0,30103	0,30103	0,30103
hasil	0	1	1	0,30103	0	0,30103
pakai	0	1	1	0,30103	0	0,30103
pengaruh	0	1	1	0,30103	0	0,30103
udah	0	2	1	0,30103	0	0,60206
lama	0	1	1	0,30103	0	0,30103
enggak	0	1	1	0,30103	0	0,30103
tau	0	1	1	0,30103	0	0,30103
tidak	0	1	1	0,30103	0	0,30103
mempan	0	1	1	0,30103	0	0,30103
senang	0	2	1	0,30103	0	0,60206
cari	0	1	1	0,30103	0	0,30103
akhir	0	1	1	0,30103	0	0,30103
obat	0	1	1	0,30103	0	0,30103
totol	0	1	1	0,30103	0	0,30103
enak	0	1	1	0,30103	0	0,30103
kering	0	1	1	0,30103	0	0,30103
kulit	1	0	1	0,30103	0,30103	0

Data *review* yang sudah melalui proses mulai dari *labelling*, pra proses, sampai pembobotan kata kemudian dilakukan klasifikasi dengan metode *Multinomial Naïve Bayes*. Namun karena data *review* yang diambil *imbalance*, dilakukan penerapan SMOTE terlebih dahulu sebelum melakukan klasifikasi. Seperti yang sudah dipaparkan di awal, terdapat 342 *review* dengan label positif dan 108 *review* dengan label negatif sehingga

terjadi ketimpangan antar kedua label. Sebelum dilakukan penerapan SMOTE, model tidak maksimal dalam mengklasifikasikan *review* negatif karena kurangnya data latih untuk *review* negatif hingga ketika dilakukan pengujian dengan *confusion matrix* hasilnya seperti pada tabel 4.

Tabel 4. Hasil Pengujian Tanpa Penerapan SMOTE

Prediksi	Aktual	
	Positif	Negatif
Positif	70 (TP)	18 (FP)
Negatif	0 (FN)	2 (TN)

Sesuai dengan tabel 4 dan perhitungan yang dilakukan dengan persamaan 8, 9, 10, dan 11, berikut merupakan hasil evaluasi dari pengujian tanpa penerapan SMOTE.

1. *Accuracy*

$$\frac{TP+TN}{Total} = \frac{70+2}{70+2+18+0} = \frac{72}{90} \times 100\% = 80\%$$
2. *Precision*

$$\frac{TP}{FP+TP} = \frac{70}{18+70} = \frac{70}{88} \times 100\% = 79\%$$
3. *Recall*

$$\frac{TP}{TP+FN} = \frac{70}{70+0} = \frac{70}{70} \times 100\% = 100\%$$
4. *Specificity*

$$\frac{TN}{TN+FP} = \frac{2}{2+18} = \frac{2}{20} \times 100\% = 10\%$$

Dapat disimpulkan bahwa nilai *specificity* atau rasio prediksi benar negatif terhadap seluruh data negatif berada di angka 10% yang artinya model ini sangat kurang dalam mendeteksi *review* negatif. Solusi dari ketimpangan ini adalah *oversampling* dengan teknik SMOTE. Teknik SMOTE mengatasi ketimpangan dengan cara memperbanyak data minoritas menggunakan data sintetik yang berasal dari kelas minoritas. Data dibagi menjadi 80% data latih dan 20% data uji, kemudian dilakukan penerapan teknik SMOTE pada data latih agar data latih positif dan negatif seimbang [12]. Data latih positif terdiri dari ± 273 *review*, yaitu 80% dari 342 data keseluruhan *review* positif dan data latih negatif terdiri dari ± 87 *review*. Setelah dilakukan penerapan SMOTE, total data latih masing-masing menjadi 273 *review* positif dan 273 *review* negatif.

Tabel 5. Pembagian Data Latih dan Uji Sebelum Penerapan SMOTE

	Positif	Negatif	Total
Data Latih	273	87	360
Data Uji	70	20	90
Total	342	108	450

Tabel 6. Hasil Data Latih Setelah Penerapan SMOTE

	Positif	Negatif	Total
Data Latih	273	273	546

Data *review* yang sudah melalui proses pembobotan kata TF-IDF dan penerapan SMOTE setelah dibagi kemudian dilakukan perhitungan *Multinomial Naïve Bayes*. Perhitungan ini menghitung probabilitas setiap kata di setiap label pada data latih. Setelah proses pada data latih selesai, selanjutnya menghitung probabilitas kondisional *term* pada data uji. Proses ini melakukan test pada data uji dengan menggunakan data latih yang telah dihitung dengan perhitungan *Multinomial Naïve Bayes*. Pada tahap ini dilakukan pengujian dengan menggunakan data uji yang mengacu model *Multinomial Naïve Bayes* yang telah dibangun pada tahap latih. Tahap uji yang mengacu pada model latih dengan menggunakan algoritma *Multinomial Naïve Bayes* akan menghitung nilai probabilitas untuk setiap kata dan menghitung nilai probabilitas kelasnya. Perhitungan dilakukan untuk setiap kemungkinan yang mungkin terjadi, baik probabilitas positif maupun negatif.

Hasil perhitungan *Multinomial Naïve Bayes* kemudian dievaluasi menggunakan beberapa pengukuran yaitu, *accuracy*, *precision*, *recall*, dan *specificity*. Namun untuk menghitung masing-masing pengukuran tersebut diperlukan tabel *confusion matrix* yang sebagai berikut:

Tabel 7. Hasil *Confusion Matrix*

Prediksi	Aktual	
	Positif	Negatif
Positif	66 (TP)	8 (FP)
Negatif	4 (FN)	12 (TN)

Nilai *true positive*, *false positive*, *true negative*, dan *false negative* yang tertera di tabel 5 kemudian dihitung dengan persamaan 8, 9, 10, dan 11.

1. *Accuracy*

$$\frac{TP+TN}{Total} = \frac{66 + 12}{66+12+4+8} \times 100\% = \frac{78}{90} \times 100\% = 86,6\%$$
2. *Precision*

$$\frac{TP}{FP + TP} = \frac{66}{8+66} = \frac{66}{74} \times 100\% = 89,1\%$$
3. *Recall*

$$\frac{TP}{TP + FN} = \frac{66}{66+4} = \frac{66}{70} \times 100\% = 94,2\%$$
4. *Specificity*

$$\frac{TN}{TN + FP} = \frac{12}{12+8} = \frac{12}{20} \times 100\% = 60\%$$

Accuracy, atau rasio prediksi benar (positif dan negatif) terhadap seluruh data, mendapatkan nilai sebesar 86,6%, dengan total 90 data yang diuji. *Precision*, atau rasio prediksi benar positif terhadap seluruh hasil yang diprediksi positif, mendapatkan nilai sebesar 89,1%. *Recall*, atau rasio prediksi benar positif terhadap seluruh data yang benar positif, mendapatkan nilai 94,2%. Terakhir, *specificity*, atau rasio prediksi benar negatif terhadap seluruh data negatif, mendapatkan nilai 60%. Dari keseluruhan hasil, dapat dilihat bahwa performa sistem masih kurang dalam memprediksi *review* negatif karena nilai *specificity*.

Tabel berikut merupakan perbandingan hasil evaluasi model sebelum dan setelah penerapan SMOTE.

Tabel 8. Perbandingan Hasil Evaluasi Model Tanpa SMOTE dan Dengan SMOTE

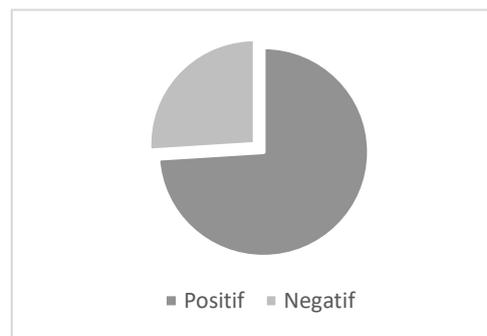
	Tanpa SMOTE	Dengan SMOTE
Accuracy	80%	86,6%
Precision	79%	89,1%
Recall	100%	94,2%
Specificity	10%	60%

Dari 450 data *review* produk Skin Game Acne Warrior, 342 *review* memiliki sentimen positif, sementara 108 *review* memiliki sentimen negatif. Berdasarkan persamaan (12) maka persentase sentimen positif dan negatif adalah sebagai berikut:

$$\text{Polaritas (kelas)} = \frac{\text{sentimen(kelas)}}{\text{total data}} \times 100\% \tag{12}$$

$$\text{Polaritas (positif)} = \frac{342}{450} \times 100\% = 76\%$$

$$\text{Polaritas (negatif)} = \frac{108}{450} \times 100\% = 24\%$$



Gambar. 2. Persentase hasil pelabelan review

Sesuai dengan gambar 2 dapat disimpulkan bahwa berdasarkan pelabelan manual, 76% user memiliki sentimen positif terhadap produk Acne Warrior, namun masih terdapat 24% user yang memiliki sentimen negatif terhadap produk Acne Warrior dari brand Skin Game.

4 Kesimpulan dan Saran

Kesimpulan yang dapat diambil dari penelitian ini adalah proses membuat model dimulai dengan pelabelan data secara manual oleh 3 penilai, kemudian dilakukan pra proses data mulai dari *case folding*, *tokenization*, normalisasi bahasa, *stopwords removal*, dan *stemming*. Proses normalisasi bahasa dan *stopwords removal* dilakukan dengan memakai *library* yang sudah ada namun juga menambahkan beberapa kata yang tidak ada di *library* sebelumnya secara manual untuk mengoptimalkan pra proses data. Selanjutnya dilakukan pembobotan kata dengan TF-IDF dan dilakukan penerapan SMOTE untuk menyeimbangkan data minoritas yaitu *review* berlabel negatif dengan data mayoritas yaitu *review* berlabel positif. Sebelum melakukan penerapan SMOTE, data dibagi dengan rasio 80:20 untuk data latih dan data uji. Data hasil pembobotan TF-IDF digunakan untuk menciptakan model klasifikasi. Tingkat akurasi dari model klasifikasi *Multinomial Naïve Bayes* pada penelitian ini sebesar 86,6%. Sedangkan untuk evaluasi dengan pengukuran lain yaitu *precision* mendapatkan nilai sebesar 89,1%, *recall* 94,2%, dan *specificity* 60%.

Berdasarkan data sebanyak 450 *review* produk Acne Warrior dari brand Skin Game di forum *review* Female Daily, 342 *review* memiliki sentimen positif dan 108 *review* memiliki sentimen negatif. Sebanyak 76% dari 450 pengguna Acne Warrior yang memberikan *review*nya di forum *review* Female Daily merasa cocok dan memiliki sentimen positif terhadap produk tersebut, sementara 24% lainnya merasa tidak cocok atau merasa produk tidak memenuhi ekspektasi. Sehingga dapat disimpulkan bahwa mayoritas dari 450 pengguna produk obat total jerawat atau *acne spot treatment* Acne Warrior yang memberikan *review* merasa produk tersebut cocok dan bisa mengatasi masalah jerawat pengguna.

Dari penelitian yang telah dilakukan, terdapat beberapa hal yang bisa diperbaiki maupun dikembangkan, Adapun saran yang diberikan untuk penelitian selanjutnya, penelitian berikutnya dapat menggunakan *annotator* profesional agar hasil pelabelan data tidak bias. Selanjutnya, dapat menggunakan lebih banyak jumlah data *review* karena *review* yang tersedia saat ini masih terbatas. Di kemudian hari bisa dilakukan pengambilan data kembali dengan kemungkinan jumlah *review* meningkat dibanding saat ini. Apabila *review* yang digunakan lebih banyak maka model dapat mempelajari lebih banyak *term* dan memberikan prediksi yang lebih baik. Terakhir, penelitian berikutnya dapat menggunakan algoritma lain seperti *Support Vector Machine* dan didukung dengan seleksi fitur.

Referensi

- [1] D. Mulyawan, dan N. Suriana, "A-Z tentang Kosmetik," Jakarta: PT Elex Media Komputindo, 2013.
- [2] N. Haasiani, "Data Penjualan Kosmetik Wajah, Brand Lokal Kuat Bersaing," diambil dari <https://compas.co.id/article/data-penjualan-kosmetik/> (diakses 2021).
- [3] B. R. Panjaitan, "4 Produk Skin Game yang Wajib Dicoba untuk Atasi Jerawat dan Komedo," <https://www.beautynesia.id/berita-skincare/4-produk-skin-game-yang-wajib-dicoba-untuk-atasi-jerawat-dan-komedo/b-189072> (diakses 2021).
- [4] Google Insight, WPP, dan Kantar. "Connected Beauty Consumer Report", 2020.
- [5] C. H. Yutika, A. Adiwijaya, dan S. Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, hal. 422-430, 2021.

- [6] A. Rahman, dan Hossen, “Sentiment Analysis on Movie Review Data Using Machine Learning Approach,” dalam *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019.
- [7] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor”, *kilat*, vol. 8, no. 1, hal. 81–92, May 2019.
- [8] A. A. Farisi, Y. Sibaroni, S. Faraby, “Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier,” *Journal of Physics: Conference Series*, vol. 1192, 2019.
- [9] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, “Tweets Classification on the Base of Sentiments for US Airline Companies,” *Entropy*, vol. 21, no. 11, hal. 1078, Nov. 2019, doi: 10.3390/e21111078.
- [10] K. Srividya, dan A. M. Sowjanya, “Aspect Based Sentiment Analysis using POS Tagging and TFIDF,” *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, hal.1960-1963. 2019, doi: 10.35940/ijeat.F7935.088619
- [11] G. P. A. Brahmantha, dan I. W. Santiyasa, “Sentiment Analysis of the Enforcement of PSBB Part II in Jakarta,” *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 9, no. 2, hal. 259-266, 2020, doi:10.24843/JLK.2020.v09.i02.p13.
- [12] W. I. Sabilla, dan C. B. Vista, “Implementation of SMOTE and Under Sampling on Imbalanced Datasets for Predicting Company Bankruptcy,” *Jurnal Komputer Terapan*, vol. 7, no. 2, 2021.