JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Analisis Sentimen Berbasis Aspek pada *Review Female Daily* Menggunakan TF-IDF dan *Naïve Bayes*

Clarisa Hasya Yutika[1], Adiwijaya[2], Said Al Faraby[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]clarisahasya@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]saidalfaraby@telkomuniversity.ac.id

Abstrak−Hasil *review* suatu produk akan memberikan manfaat yang cukup besar bagi produsen atau konsumen. Female daily merupakan salah satu forum yang membahas tentang produk kecantikan. Terdapat banyak *review* setiap harinya yang diperoleh. Maka dari itu diperlukannya teknik untuk menganalisis hasil *review* tersebut menjadi sebuah informasi yang berharga. Salah satu teknik nya adalah analisis sentimen berbasis aspek. Analisis sentimen berbasis aspek akan menganalisis setiap teks untuk mengidentifikasi berbagai aspek (atribut atau komponen) kemudian menentukan tingkat sentimen (positif, negatif, atau netral) yang sesuai untuk masing-masing aspek. Dari hasil *review* yang didapat, terdapat *review* yang menggunakan bahasa multilingual. Maka tahapan yang dilakukan adalah dengan menerjemahkan bahasa multilingual tersebut menjadi satu bahasa saja, yaitu Bahasa Indonesia. Sebelum *review* tersebut diolah, akan dilakukannya *preprocessing* supaya lebih mudah diproses. Kemudian dilakukannya pembobotan kata menggunakan TF-IDF, dan metode untuk mengklasifikasi sentimen yang akan digunakan adalah Complement Naïve Bayes untuk mengatasi data yang tidak seimbang. Dari hasil pengujian diperoleh nilai *F1-Score* sebesar 62,81% untuk data yang diterjemahkan ke dalam Bahasa Inggris kemudian ke dalam Bahasa Indonesia dan tidak menggunakan *stopword removal*.

Kata Kunci: *review female daily*, analisis sentimen berbasis aspek, *preprocessing*, TF-IDF, *Complement Naïve Bayes*

Abstract−The results of a product review will provide considerable benefits for producers or consumers. Female daily is a forum that discusses beauty products. There are many reviews that are obtained every day. Therefore a technique is needed to analyze the results of the review into valuable information. One of the techniques is aspect-based sentiment analysis. Aspect-based sentiment analysis will analyze each text to identify various aspects (attributes or components) then determine the level of sentiment (positive, negative, or neutral) that is appropriate for each aspect. From the results obtained, there are reviews that use multilingual languages. Then the steps taken are to translate the multilingual language into one language only, namely Indonesian. Before the review is processed, preprocessing will be carried out to make it easier to process. Then the word weighting is done using TF-IDF, and the method for classifying sentiments that will be used is Complement Naïve Bayes to overcome unbalanced data. From the test results obtained the best F1-Score of 62,81% for data translated into English and then into Indonesian and not using stopword removal.

Keywords: review female daily, aspect-based sentiment analysis, preprocessing, TF-IDF, Complement Naïve Bayes.

1. PENDAHULUAN

Female Daily merupakan salah satu forum yang diciptakan untuk membahas produk kecantikan. Setiap harinya banyak pengguna internet yang memberi *review* pada beberapa produk, sehingga jumlah *review* bisa berkisar dari ratusan hingga ribuan dan berisi berbagai pendapat. Dari hasil *review* tersebut dapat memberikan manfaat yang cukup besar untuk produsen dan konsumen [1]. Maka dari itu, diperlukannya mengolah hasil *review* tersebut untuk mendapatkan informasi yang lebih berharga. Salah satu caranya adalah analisis sentimen berbasis aspek, yaitu mengklasifikasikan menjadi kelas positif, netral, dan negatif. Selanjutnya akan dicari berdasarkan aspek atau dapat disebut *opinions mining* dan *opinions summarization* [2].

Penelitian mengenai analisis sentimen berbasis aspek sudah banyak dilakukan, salah satunya adalah pada tahun 2017 oleh Mubarok Et Al [1] membahas analisis sentimen berbasis aspek menggubakan metode *Naïve Bayes* yang menghasilkan performansi untuk *aspect-based sentiment analysis* menghasilkan *F1-Measure* sebesar 78.12%, kemudian untuk *aspect classification* menghasilkan *F1-Measure* sebesar 88.13%, dan untuk *sentiment classification* menghasilkan *F1-Measure* sebesar 75%. Pada penelitian ini juga menggunakan *POS tagging* dan *Chi Square*, dan terbukti bahwa *Chi Square* mempercepat waktu perhitungan dalam proses klasifikasi namun menurunkan kinerja sistem.

Analisis sentimen pada komentar *review* produk kosmetik telah dilakukan untuk menganalisis kepuasan telah dilakukan oleh Pugsee Et Al [4] menggunakan *dataset review* produk kosmetik dari www.makeupalley.com dan dilabeli secara manual dengan label positif sebanyak 2.724 komentar dan negatif sebanyak 484 *review*. Dilakukan empat tahap untuk mengaplikasikan analisis kepuasan yaitu, *POS tagging*, memeriksa kata-kata dalam leksikon sentimen kosmetik, memilih kata-kata sentimen untuk membuat model klasifikasi, dan mengklasifikasikan komentar ulasan produk menggunakan *Naïve Bayes Classifier*. Terdapat tiga skenario dalam penelitian ini, 1) semua *review* produk sentimen, 2) beberapa *review* positif dan semua *review* negatif, dan 3) semua *review* positif dan duplikat *review* negatif. Pada skenario pertama dihasilkan akurasi sebesar 83.42%, tetapi *precision* dan *recall* label negatif sangat rendah. Pada skenario kedua dihasilkan akurasi sebesar 74.79%, *precision* label negatif dan *recall* label positif cukup tinggi. Pada skenario ketiga dihasilkan akurasi sebesar 94.17%, *precision* dan *recall* semua label cukup tinggi.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Kristiyanti [5] mengklasifikasikan *review* produk kosmetik dengan label positif dan negatif menggunakan algoritma *Support Vector Machine* (SVM) dan *Particle Swarm Optimization* untuk seleksi fitur. Perbedaan akurasi sebelum seleksi fitur dan sesudah cukup meningkat, yaitu dari 89.0% menjadi 97.0%.

Penelitian selanjutnya pada tahun 2019 oleh Srividya Et Al [6] melakukan analisis sentimen berbasis aspek menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Kemudian terdapat dua model, yaitu model 1 menggunakan *POS tagging*, dan model 2 menggunakan TF-IDF. Hasil penelitian ini menujukan bahwa performansi pada model 2 lebih bagus dibandingkan dengan model 1, kemudian performansi *Support Vector Machine* (SVM) lebih tinggi dibandingkan *Naïve Bayes.*

Penelitian selanjutnya oleh Gojali Et Al [3] menggunakan *dataset review* restoran yang berbahasa *multilingual* dan menggunakan algoritma *Naïve Bayes Classifier* untuk *subjectivity classification* dengan menghasilkan F1-Measure sebesar 78.3% dan menjadi model yang terbaik, dan *Conditional Random Field* (CRF) dikombinasikan dengan *lexical* dan *POS tagging* menghasilkan *F1-Measure* sebesar 79.4% dan menjadi fitur yang terbaik.

Kemudian penelitian tentang klasifikasi teks oleh Xhemali Et Al [7] membandingkan tiga metode, yaitu *Naïve Bayes*, *Decision Trees*, dan *Neural Network*, menunjukan bahwa nilai performansi pada *Naïve Bayes* adalah 95.20% *accuracy* dan 97.26% F-Measure. Penelitian serupa juga dilakukan oleh Dong Et Al [8] dengan membandingkan tiga metode, yaitu *k-Nearest Neighbor*, *Naïve Bayes*, dan *Improved Naïve Bayes* yang menggunkan *Gini Index* dan TF-TWF, menunjukan bahwa performansi *Improved Naïve Bayes* naik dari 10 ke 20 persen.

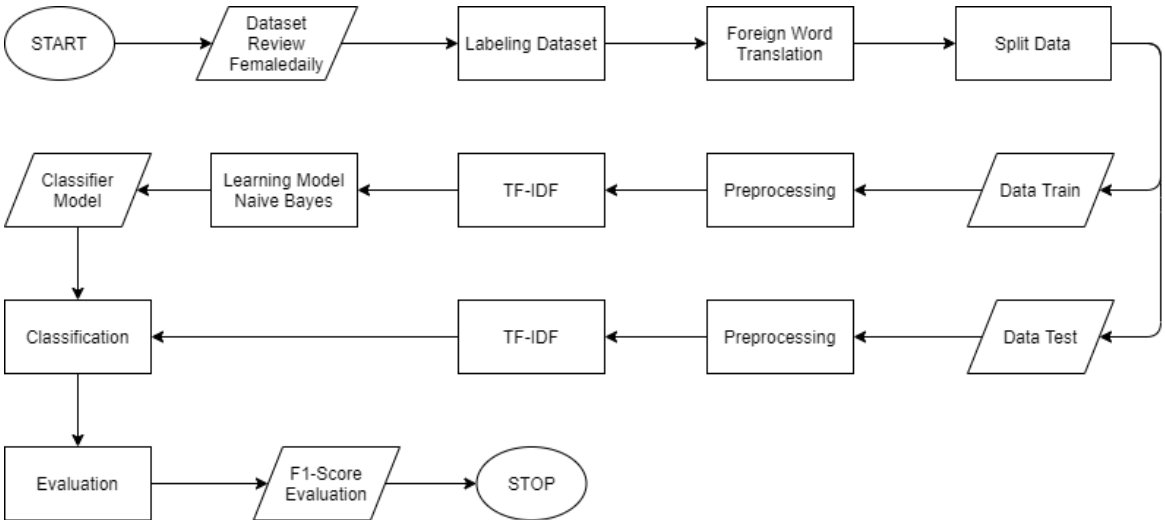
Sehingga pada penelitian ini bertujuan untuk menganalisis sentimen berbasis aspek menggunakan *dataset review female daily* yang berbahasa *multilingual* dengan menggunakan dan tidak menggunakan tahapan *foreign word translation* untuk mengatasi *dataset* yang berbahasa multilingual, dan pengaruh *preprocessing stopword* removal dan *stemming*, dengan menggunakan pembobotan TF-IDF dan algoritma *Naïve Bayes*. Untuk menghitung performansinya menggunakan metode evaluasi *accuracy*, *precision*, *recall*, dan *F1-Score*.

Batasan pada penelitian ini adalah *dataset* yang digunakan bersumber dari *web female daily*, dan hanya mengambil kategori *serum*, *toner*, *sunscreen*, *scrub*, dan *exfoliator*. *Dataset* berjumlah 5054 *review*, dengan 4 aspek label yaitu harga, kemasan, produk, aroma dengan kelas positif, netral, dan negatif. Pelabelan dilakukan secara manual oleh 4 orang.

2. METODOLOGI PENELITIAN

2.1 Rancangan Sistem

Pada tahap ini akan dijelaskan tentang sistem yang dibangun. Berikut adalah flowchart yang menggambarkan alur kerja pembangunan sistem secara umum pada Gambar 1.



Gambar 1. Gambaran Umum Perancangan Sistem

2.2 *Dataset*

Penelitian ini menggunakan 5054 *review* yang didapat dari web female daily. Kategori produk yang digunakan adalah *toner*, *serum*, *sunscreen*, *scrub*, dan *exfoliator*. Contoh data female daily yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Tabel 1. Contoh *dataset review* female daily dengan kategori produk

Review	Kategori Produk
wonder pore ini lumayan agak ngefek sih di aku, biasa pake ini buat toner. tapi untuk serinya wonder pore, kayaknya wonder pore ini agak biasa aja untuk ngecilin pori pori, tapi wonder pore freshner ini punya terobosan buat ngebasmi kuman dan cacing yang ada di kulit kita supaya kulit kita ngga berpori besar dan berminyak	Toner
love it so much! awalnha aku bukan orang yg suka pakai sunblock karena aku gak suka sama efek lengketnya. namun sejak ketemu sama biore ini, aku jadi sukaa banget krn sublock ini sama sekali gak lengket di kulit. wanginya seger kayak ada citrusnya gitu and adem di wajah. selain itu juga sangat cepat meresap ke kulit wajah karena sifatnya yang water base	Sunscreen
Pros wanginya enak scrub nya oke ga terlalu kasar cukup buat ngangkat kotoran dan bikin kulit halus harga terjangkau packaging okelah so so sih tapi dengan harga segini cukup kepake lama sih cons - one of my fave repurchase: yes	Scrub & Exfoliator

2.3 *Labelling Dataset*

Dataset dilabeli secara manual oleh 4 orang, setiap orang melabeli satu kategori produk. Tahap pertama pelabelan adalah identifikasi aspek, terdapat 4 aspek yaitu harga, kemasan, produk, dan aroma. Dari setiap aspek terdapat beberapa kata yang menunjukan bahwa *review* tersebut membahas aspek yang dituju. Identifikasi aspek dan kata dijelaskan pada Tabel 2.

Tabel 2. Identifikasi aspek kategori

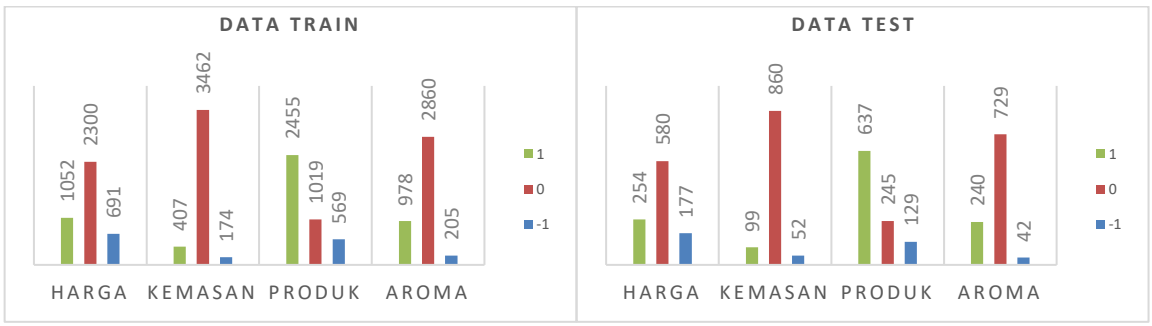
Aspek Kategori	Kata
Harga	Mahal, Murah, Diskon, Terjangkau
Kemasan	Praktis
Produk	Favorit, Cocok, Suka
Aroma	Wangi, Bau, Menyengat

Kemudian aspek yang dibahas dalam suatu *review* akan diklasifikasikan menjadi 3 kelas polarity, yaitu kelas positif diberi nilai “1”, kelas netral diberi nilai “0”, dan kelas negatif diberi nilai “-1”. Jika terdapat aspek yang tidak dibahas dalam suatu *review*, maka akan diberi nilai “0”. Contoh pelabelan dapat dilihat pada Tabel 3.

Tabel 3. Contoh pelabelan *dataset*

Review	Harga	Kemasan	Produk	Aroma
Pros wanginya enak scrub nya oke ga terlalu kasar cukup buat ngangkat kotoran dan bikin kulit halus harga terjangkau packaging okelah so so sih tapi dengan harga segini cukup kepake lama sih cons - one of my fave repurchase: yes	1	0	1	1

Berdasarkan identifikasi aspek yang telah dilakukan, *review* pada tabel 3 masuk ke dalam 3 aspek, yaitu aspek harga karena didalam *review* terdapat kata ‘harga terjangkau’, aspek produk karena didalam *review* terdapat kata ‘fave’ yang artinya favorit, aspek aroma karena didalam *review* terdapat kata ‘wanginya enak’. Untuk aspek kemasan diberi nilai’0’ karena didalam *review* hanya membahas ‘packaging okelah so so’ yang artinya kemasan tersebut biasa-biasa saja.
Setelah *dataset* selesai dilabeli, *dataset* akan di split menjadi 80% data train sejumlah 4043 dan 20% data test sejumlah 1011. Berikut adalah jumlah kelas setiap polarity dalam setiap aspek dapat dilihat pada Gambar 2.



Gambar 2. Data Statistik Kelas Polarity Dalam Setiap Aspek

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

2.4 *Foreign word translation*

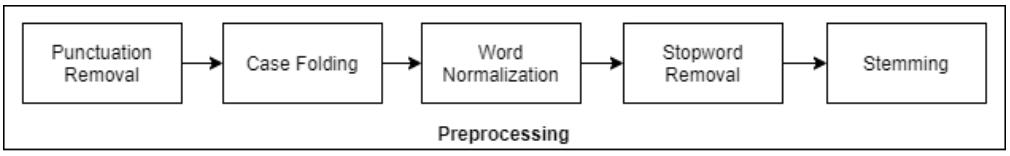
Pada tahap ini, dilakukan penerjemahan kata asing dalam kalimat ke dalam Bahasa Indonesia [3]. Terdapat 2 skenario dalam penerjemahan, skenario pertama seluruh *dataset* diterjemahkan ke Bahasa Indonesia, skenario kedua adalah seluruh *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia. Tujuan dilakukannya skenario kedua adalah, pada saat proses penerjemahan ke Bahasa Indonesia masih terdapat beberapa kata yang tidak diterjemah. Sehingga penulis melakukan eksperimen untuk menerjemahkan ke dalam Bahasa Inggris terlebih dahulu kemudian diterjemahkan lagi ke Bahasa Indonesia. Proses penerjemahan menggunakan google translate. Adapun ilustrasi penerjemahan dapat dilihat pada Tabel 4.

Tabel 4. Penerjemahan *dataset*

Dataset	*Review*
Non Translate	pros wanginya enak scrub nya oke ga terlalu kasar cukup buat ngangkat kotoran dan bikin kulit halus harga terjangkau packaging okelah so so sih tapi dengan harga segini cukup kepake lama sih cons - one of my fave repurchase: yes
Translate ID	pro wanginya enak scrub nya oke ga terlalu kasar cukup buat kotoran kotoran dan kulit halus harga terjangkau packaging okelah so so sih tapi dengan harga segini cukup kepake lama sih kontra - one of my fave repurchase: yes
Translate EN - ID	pro wangi lulur oke tidak terlalu kasar untuk menghilangkan kotoran dan menghaluskan kulit, harga terjangkau, kemasan oke, tapi dengan harga ini butuh waktu yang lumayan lama kontra - salah satu favorit saya untuk membeli kembali: ya

2.5 *Preprocessing*

Tahap ini merupakan tahapan paling penting sebelum melakukan klasifikasi, karena terdapat banyak . Terdapat beberapa tahapan dalam proses *preprocessing* yang dapai dilihat pada Gambar 3.



Gambar 3. Tahapan *Preprocessing*

2.5.1 *Punctuation Removal*

Pada tahap ini, karakter selain huruf dihilangkan dan dianggap delimiter atau dihapus [9], termasuk tanda baca, dan angka.

2.5.2 *Case Folding*

Pada tahap ini mengubah semua huruf dalam teks menjadi huruf kecil.

2.5.3 *Word Normalization*

Pada tahap ini, dilakukan proses *no*rmalisasi kata yaitu mengubah kata tidak baku menjadi baku. Kamus *no*rmalisasi kata didapat dari data train, dan dibuat oleh 4 orang. Terdapat 387 kata dalam kamus.

2.5.4 *Stopword Removal*

Pada tahap ini, dilakukan penghapusan kata-kata yang dianggap tidak sesuai atau sering muncul seperti: ‘di’, ‘ke’, ‘dari’, ‘yang’, ‘dan’, ‘atau’, ‘ini’, dan lainnya. Proses ini menggunakan list *stopword* dari Sastrawi dengan menghapus kata negasi seperti ‘tidak’ dan ‘nggak’. Kemudian ditambahkan 142 kata yang tidak memiliki arti seperti ‘argh’, ‘ooh’, ‘haha’, dan lainnya.

2.5.5 *Stemming*

Pada tahap ini, dilakukan membersihkan data dari kata imbuhan, awalan, sapaan, akhiran, atapun kombinasi. Dengan demikian setiap kata pada dokumen hanya mengandung kata dasar saja. Proses ini menggunakan *library* Sastrawi.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Tabel 5. Hasil *Preprocessing*

Proses	Hasil Proses
Data	Wanginya enak scrub nya oke ga terlalu kasar, dan bikin kulit halus. Harga terjangkau.
Punctuation Removal	Wanginya enak scrub nya oke ga terlalu kasar dan bikin kulit halus Harga terjangkau
Case Folding	wanginya enak scrub nya oke ga terlalu kasar dan bikin kulit halus harga terjangkau
Word Normalization	wanginya enak scrub nya ok tidak terlalu kasar dan membuat kulit halus harga terjangkau
Stopword Removal	wanginya enak scrub tidak terlalu kasar membuat kulit halus harga terjangkau
Stemming	wangi enak scrub tidak terlalu kasar kotor buat kulit halus harga jangkau

2.6 TF-IDF

TF-IDF merupakan fitur pembobotan paling populer yang sering digunakan serta memiliki akurasi dan *recall* yang cukup tinggi [10]. Pembobotan TF-IDF adalah jenis pembobotan yang sering digunakan dalam *information retrieval.* Pembobotan TF-IDF dinilai penting, apabila sebuah kata lebih sering muncul dalam sebuah dokumen maka nilai kontribusinya akan semakin besar, namun apabila kata tersebut sering muncul dalam beberapa dokumen maka akan memiliki kontribusi yang lebih kecil [11]. TF-IDF terdiri atas *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).
Term frequency menyatakan nilai frekuensi *term* yang sering muncul pada sebuah dokumen. Semakin besar jumlah kemunculan suatu *term* dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar. $f(t_k, d_j)$ mendefinisikan jumlah kemunculan term *k* pada sebuah dokumen *j*. Rumus TF dapat didefinisikan dengan [12]:

$$TF(t_k, d_j) = f(t_k, d_j) \quad (1)$$

Pada sebuah koleksi dokumen *term* dapat didistribusikan secara acak dengan menggunakan IDF. Semakin sering suatu *term* muncul di banyak dokumen maka nilai IDFnya akan kecil. Jumlah dokumen pada *dataset* (D) akan dibagi dengan jumlah dokumen yang mengandung *term* $df(t)$. IDF dapat dihitung dengan rumus sebagai berikut [12] :

$$IDF(t_k) = \log \frac{D}{df(t)} \quad (2)$$

TF-IDF dapat dirumuskan sebagai berikut [12] :

$$TF\ IDF(t_k, d_j) = TF(t_k, d_j) * IDF(t_k) \quad (3)$$

Untuk mengetahui peranan fitur dalam setiap dokumen dan kategori kelas yang ada. ‘D1’, ‘D2’, ‘D3’ , ‘D4’ dan ‘D5’ menunjukan dokumen yang ada. Adapun ilustrasi perhitungan TF-IDF dapat dilihat pada Tabel 6.

Tabel 6. Perhitungan TF-IDF

Dokumen	TF					DF	IDF	TF.IDF				
Term	D1	D2	D3	D4	D5			D1	D2	D3	D4	D5
Cocok	1	0	1	0	1	3	log(5/3)=0.2	0.2	0	0.2	0	0.2
Tekstur	1	1	1	0	0	3	log(5/3)=0.2	0.2	0.2	0.2	0	0
Kulit	0	0	1	1	0	2	log(5/2)=0.4	0	0	0.4	0.4	0
Produk	0	1	0	2	0	3	log(5/3)=0.2	0	0.2	0	0.4	0
Harga	1	1	0	0	1	3	log(5/3)=0.2	0.2	0.2	0	0	0.2

2.7 *Naïve Bayes Classifier*

Naïve Bayes Classifier merupakan model probabilistik yang digunakan untuk proses klasifikasi berdasarkan teorema Bayes. *Naïve Bayes Classifier* dikenal sederhana namun sangat efisien [13], dan memiliki asumsi yg sangat kuat terhadap independensi dari masing-masing kondisi, terutama saat memiliki jumlah data train yang sedikit. Pada penelitian ini menggunakan algoritma *Complement Naïve Baye*s yang merupakan adaptasi dari algoritma *Multinomial Naïve Bayes* dan dirancang untuk mengatasi *dataset* yang tidak seimbang [14]. Data yang tidak seimbang adalah ketika jumlah suatu kelas lebih tinggi dibandingkan kelas lainnya, sehingga dapat memengaruhi perhitungan probabilitas kelas lain.
Complement Naïve Bayes bekerja dengan menghitung probabilitas kata yang muncul selain dalam kelasnya. Kemudian dihitung probabilitas setiap kelas dan dipilih nilai probabilitas terendah. Nilai probabilitas terendah dipilih karena nilai tersebut bukan dihasilkan dari kelas tersebut. Sehingga menyiratkan bahwa kelas tersebut memiliki probabilitas tertinggi untuk kelas tersebut.

$$P(c) = \frac{N_c}{N} \quad (4)$$

$$P(w|\hat{c}) = \frac{count(\hat{c},w)+\alpha_w}{count(\hat{c})+\alpha} \quad (5)$$

$$P(c|w_i) = argmin\ P(c) \prod_{i=1} \frac{1}{P(w|\hat{c})^{fi}} \quad (6)$$

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Dari persamaan (4), (5) dan (6), w adalah kata, dan c adalah kelas target. N_c adalah jumlah kelas c. N adalah jumlah seluruh kelas. $Count(\hat{c}, w)$ adalah berapa kali kata w muncul selain dalam dokumen selain kelas c. $Count(\hat{c})$ adalah jumlah seluruh kata yang muncul di kelas selain c. α_w adalah *smoothing parameter*, biasanya diberi nilai 1. α adalah jumlah keseluruhan dari α_w. f_i adalah jumlah frekuensi kata i dalam dokumen w_i. $P(w|\hat{c})$ dapat disebut sebagai *likelihood* yang dapat diartikan probabilitas dari kata w yang muncul selain dalam kelas c. $P(c)$ biasa disebut *prior probability* yang dapat diartikan probabilitas kelas c yang muncul dalam seluruh kelas. $P(c|w_i)$ biasa disebut *posterior probability* yang dapat diartikan nilai terendah dari probabilitas kelas c yang muncul dalam dokumen w_i.

2.8 Evaluasi

Evaluasi sistem dilakukan untuk mengetahui seberapa baik performansi sistem yang dihasilkan. Metode evaluasi yang akan digunakan adalah *accuracy*, *precision*, *recall*, dan *F1-Score*. Kemudian nilai evaluasi yang akan dibandingkan hanya nilai *F1-Score* saja, dikarenakan *F1-Score* adalah metrik yang lebih baik untuk data yang tidak seimbang dibandingkan dengan *accuracy*. Untuk menghitung metode evaluasi yang sudah disebutkan, dibutuhkan *True Prositive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dari *confusion matrix* setiap label.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (7) \qquad precision = \frac{TP}{TP+FP} \quad (8) \qquad recall = \frac{TP}{TP+FN} \quad (9) \qquad F1\ Score = \frac{2\times(recall\ \times precision)}{(recall+precision)} \quad (10)$$

Contoh perhitungan untuk salah satu label dapat dilihat pada Tabel 7.

Tabel 7. Perhitungan *Accuracy*, *Precision*, *Recall*, *F1-Score* untuk 1 label

Kelas	TP	FP	TN	FN	*Accuracy*	*Precision*	*Recall*	*F1-Score*
1	100	10	50	5	0,94	0,91	0,95	0,93
0	98	6	54	7	0,95	0,94	0,93	0,94
-1	103	4	46	2	0,93	0,96	0,98	0,97
Macro Average					0,94	0,94	0,95	0,94

Setelah didapat hasil macro average setiap label, kemudian dihitung rata-rata untuk semua label yang dapat dilihat pada Tabel 8.

Tabel 8. Perhitungan A, P, R, F1 untuk multi-label

Aspek	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Harga	75,27%	74,19%	68,35%	69,29%
Kemasan	85,76%	52,84%	43,45%	45,48%
Produk	70,92%	66,28%	59,26%	61,29%
Aroma	78,14%	48,21%	49,35%	48,75%
Rata-rata	77,52%	60,38%	55,10%	56,20%

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan *dataset review* female daily sebanyak 5054 yang dibagi menjadi 80% data train sejumlah 4043 dan 20% data test sejumlah 1011. Setiap data memiliki 4 label aspek yang masing-masing terdapat 3 kelas polarity, untuk jumlah setiap kelas polarity dapat dilihat pada Gambar 2. Terdapat beberapa skenario pengujian. Skenario pertama adalah membandingkan *dataset* yang belum di terjemahkan dan *dataset* yang sudah diterjemahkan. Kemudian skenario kedua membandingkan hasil yang menggunakan *preprocessing stopword* dan *stemming*, tanpa *stopword*, tanpa *stemming*, dan tanpa *stopword* dan *stemming*. Kemudian skenario ketiga dilakukan *hyperparameter tuning* untuk mendapatkan parameter terbaik.
Tujuan dilakukannya ketiga skenario tersebut adalah untuk melihat performansi *dataset* yang tidak diterjemahkan dan diterjemahkan serta pengaruh *preprocessing stopword* dan *stemming* menggunakan pembobotan TF-IDF dan algoritma Naïve Bayes.

3.1. Evaluasi Skenario I

Pada pengujian skenario I dilakukan pengujian dengan *dataset* yang belum diterjemahkan, *dataset* diterjemahkan ke Bahasa Indonesia, dan *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia. Alasan dataset diterjemahkan adalah pada beberapa *review* terdapat penggunaan kata dalam Bahasa Inggris.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

Tabel 9. Hasil Skenario I

Skenario	*Accuracy*	*Precision*	*Recall*	*F1-Score*
*N*on Translate	75,77%	62,18%	65,60%	62,34%
Translate ID	75,20%	62,60%	65,34%	62,45%
Translate EN-ID	75,99%	63,21%	66,03%	62,81%

Berdasarkan hasil skenario I, dihasilkan *F1-Score* tertinggi sebesar 62,81% untuk *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia. Dapat disimpulkan proses *foreign word translation* dapat meningkatkan performansi.

3.2 Evaluasi Skenario II

Pada pengujian skenario I didapatkan dataset terbaik yaitu *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia. Sehingga untuk skenario II menggunakan dataset tersebut.

Tabel 10. Hasil Skenario II

Skenario	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Full Preprocessing	75,57%	62,54%	65,79%	62,31%
No Stopword Removal	75,99%	63,21%	66,03%	62,81%
No Stemming	75,87%	62,84%	65,48&	62,27%
No Stopword No Stemming	75,49%	60,95%	62,91%	60,63%

Hasil dari skenario II, dihasilkan *F1-Score* tertinggi sebesar 62,81% untuk *dataset* yang tidak menggunakan tahapan *stopword removal*. Dapat disimpulkan bahwa ketika menggunakan tahapan *stopword removal*, terdapat kata-kata yang dapat menambah informasi dalam kalimat tersebut dihilangkan, sehingga dapat mengubah makna kalimat yang seharusnya. Dapat dilihat pada Tabel 11, pada *review* yang menggunakan semua tahapan *preprocessing*, untuk aspek kemasan terdapat kesalahan prediksi. Pada kalimat 'packagingnya oke' yang artinya kemasan tersebut bagus, tetapi ketika kata 'ok' dihilangkan pada tahapan *stopword removal* hanya tersisa kata 'kemas' saja yang tidak memiliki makna, sehingga diklasifikasikan menjadi kelas netral.

Tabel 11. Kesalahan Klasifikasi

Data	Review	Harga	Kemasan
Asli	sunscreen ini nggak bikin wajah abu-abu, harganya juga terjangkau dan packagingnya oke.	1	1
Full Preprocessing	sunscreen tidak buat wajah abu abu harga jangkau kemas	1	0
No Stopword Removal	sunscreen ini tidak buat wajah abu abu harga juga jangkau dan kemas ok	1	1

3.3 Evaluasi Skenario III

Pada pengujian skenario sebelumya didapatkan dataset terbaik yaitu *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia dan tidak menggunakan tahapan *preprocessing stopword removal*. Kemudian dilakukan *hyperparameter tuning* menggunakan metode *Grid Search* sebanyak *10-Fold cross validation* pada *data train,* sehingga terbagi menjadi *data validation* dan *data train*. Parameter yang digunakan merupakan hasil dari parameter terbaik pada *data validation*. Proses ini dilakukan untuk mencari parameter terbaik. Parameter yang di *tune* adalah 'min_df', 'max_df', 'max_features' yang merupakan parameter dari TF-IDF, dan 'alpha' yang merupakan parameter dari ComplementNB. 'min_df' digunakan untuk menghilangkan kata yang jarang muncul, misal 'min_df = 0.01' artinya akan dihapus kata-kata yang muncul kurang dari 1% dari dokumen, atau 'min_df = 5' artinya akan dihapus kata-kata yang hanya muncul kurang dari 5 dokumen. 'max_df' digunakan untuk menghapus kata yang sering muncul, misal 'max_df = 50' artinya akan dihapus kata-kata yang muncul lebih dari 50% dokumen, atau 'max_df = 50' artinya akan dihapus kata-kata yang muncul lebih dari 50 dokumen. 'max_features' digunakan untuk membatasi berapa banyak kata-kata yang akan digunakan, misal 'max_features = 4000' maka membuat feature matrix dari 4000 kata yang sering muncul. 'alpha' adalah parameter untuk smoothing. Parameter ini merupakan parameter dari *library* scikit-learn TfidfVectorizer dan ComplementNB [15].

Tabel 12. Hasil Hyperparameter Tuning

CNB alpha	TFIDF min_df	TFIDF max_df	TFIDF max_features	*F1-Score*
1	0,01	0,7	2000	61,86%
0,7	3	1,0	4000	60,32%

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

0,4	2	0,9	2000	61,73%
0,4	0,001	0,5	4000	61,49%

Dari Tabel 12 didapatkan parameter terbaik dengan *F1-Score* sebesar 61,86%. Sehingga pada *data test* menggunakan paramater tersebut.

Tabel 13. Hasil Skenario III

Aspek	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Harga	79,92%	78,50%	80,05%	78.24%
Kemasan	79,23%	52,75%	56,72%	51,25%
Produk	67,95%	63,06%	68,79%	64,33%
Aroma	76,85%	58,53%	58,56%	57,41%
Rata-rata	75,99%	63,21%	66,03%	62,81%

Setelah menggunakan parameter terbaik, didapatkan *F1-Score* sebesar 62,81%. Berdasarkan Tabel 13, aspek harga mendapatkan performansi yang paling tinggi dibanding aspek lainnya. Hal ini disebabkan karena distribusi data pada aspek harga lebih seimbang, dan identifikasi kata untuk pelabelan aspek lebih spesifik dibanding ketiga aspek lainnya. Pada aspek kemasan dan aroma kelas netral menjadi kelas mayor, dikarenakan banyak *review* yang tidak membahas kemasan dan aroma sehingga dilabeli netral. Sehingga terjadinya data yang tidak seimbang, mengakibatkan model tidak dapat memprediksi dengan baik. Meskipun aspek produk distribusi datanya hampir seimbang, tetapi performansinya tidak cukup baik. Hal tersebut terjadi dikarenakan proses pelabelan yang tidak konsisten, identifikasi kata aspek produk kurang spesifik karena aspek produk diberikan label berdasarkan opini secara luas sehingga sulit untuk melabeli aspek produk. Misal terdapat kalimat 'aku suka produk ini, tapi aku gamau beli lagi soalnya susah dicari' yang memiliki sentimen positif dan negatif. Kemudian pelabelan ini dilakukan secara manual oleh 4 orang, dan bisa saja terjadi pendapat yang berbeda setiap orang.
Perbandingan nilai *accuracy* dengan precision, recall dan *F1-Score* pada setiap aspek terdapat selisih yang cukup signifikan. Hal ini dapat terjadi karena *accuracy* hanya fokus kepada hasil prediksi yang benar (*True Positive* dan *True Negative*). Kemudian nilai *recall* lebih tinggi dibanding *precision*. Hal ini terjadi karena model lebih baik mengenali data positif dibandingkan dengan keseluruhan data positif, dan lebih baik menghindari terjadinya *False Positive*. Nilai precision rendah yang artinya model sulit untuk memprediksi data yang positif karena data yang tidak seimbang, sehingga model cenderung memprediksi kelas yang jumlahnya lebih banyak. Dikarenakan nilai *precision* dan *recall* yang rendah, mengakibatkan nilai *F1-Score* yang rendah.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa data yang tidak seimbang dapat memengaruhi performansi. Dari keempat aspek, distribusi data pada aspek kemasan, produk dan aroma sangat tidak seimbang, sehingga performansi menjadi rendah. Sedangkan untuk performansi pada aspek harga yang paling tinggi dibanding yang lain, karena distribusi datanya yang cukup seimbang dibanding aspek lainnya. Kemudian untuk data yang diterjemah dapat meningkatkan performansi, karena jika data belum diterjemah akan terdapat dua kata yang memiliki satu makna, dan ketika sudah diterjemah kata tersebut menjadi sama. Tahapan *preprocessing* yang tidak menggunakan *stopword removal* mendapat performansi yang paling tinggi, hal ini dapat terjadi karena terdapat kata-kata yang dapat menentukan suatu sentimen kalimat tidak dihilangkan. Dari ketiga skenario yang telah dilakukan, didapat performansi tertinggi oleh *dataset* diterjemahkan ke dalam Bahasa Inggris kemudian diterjemahkan ke Bahasa Indonesia dan tidak menggunakan *stopword removal* dengan parameter alpha atau smoothing sebesar 1, min_df sebesar 0,01, max_df sebesar 0,7, dan max_features sebesar 2000 menghasilkan performansi terbaik dengan nilai *F1-Score* sebesar 62,81%.
Saran untuk penelitian selanjutnya adalah melakukan analisis sentimen berbasis ontologi dan mencoba teknik terjemahan selain dari *google translate*, serta memperbaiki kamus kata untuk *stopword removal*

REFERENCES

[1] Mubarok, M. S., Adiwijaya, & Aldhi, M. D. (2017, August). Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes. *AIP Conference Proceedings. Vol. 1867, No. 1, p. 020060.* AIP Publishing LLC.
[2] Hu, M., & Liu, B. (2004, August). Mining and Summarizing Customer *Reviews*. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 168-177).
[3] Gojali, S., & Khodra, M. L. (2016, August). Aspect Based Sentiment Analysis for Review Rating Prediction. *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)* (pp. 1-6). IEEE.
[4] Pugsee, P., Sombatsri, P., & Juntiwakul, R. (2017, May). Satisfactory analysis for cosmetic product review comments. *Proceedings of the 2017 International Conference on Data Mining, Communications and Information Technology* (pp. 1-6).
[5] Kristiyanti, D. A. (2015). Analisis Sentimen *Review* Produk Kosmetik menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization sebagai Metode Seleksi Fitur. *SNIT 2015, 1(1)*, 134-141.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor X, Bulan 2021, Page 999-999
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2293

[6] Srividya, K., & Sowjanya, A. M. (2019). Aspect Based Sentiment Analysis using POS Tagging and TFIDF. *International Journal of Engineering and Advanced Technology (IJEAT). Volume-8 Issue-6.* Blue Eyes Intelligence Engineering & Sciences Publication.

[7] Xhemali, D., Hindie, C. J., & Stone, R. G. (2009, September). Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *International Journal of Computer Science Issues (IJCSI), Volume 4, Issue 1*, pp. 16-23.

[8] Dong, T., Shang, W., & Zhu, H. (2011). An Improved Algorithm of Bayesian Text Categorization. *JSW, Volume 6. Issue 9*, pp. 1837-1843.

[9] Uysal, A. K., & Gunal, S. (2014). The Impact of Preprocessing on Text Classification. *Information Processing and Management, Vol. 50*, pp. 104-112.

[10] Ye, J., Jing, X., & Li, J. (2017, September). Sentiment Analysis Using Modified LDA. *International conference on signal and information processing, networking and computers* (pp. 205-212). Springer, Singapore.

[11] Yulietha, I. M., Faraby, S. A., & Adiwijaya. (2017). Klasifikasi Sentimen *Review* Film Menggunakan Algoritma Support Vector Machine Sentiment Classification of Movie *Review*s Using Algorithm Support Vector Machine. *eProceedings of Engineering, Volume 4. Issues 3.*

[12] Nugraha, M. (2014). Sentimen Analysis *Review* Film dengan menggunakan metode KNN. Bandung: Widyatama University.

[13] Raschka, S. (2014). *Naive bayes* and Text Classification I - Introduction and Theory. *arXiv preprint.* Retrieved from https://arxiv.org/abs/1410.5329

[14] Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of *naive bayes* text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.

Submitted: **dd/mm/yyyy**; Accepted: **dd/mm/yyyy**; Published: **dd/mm/yyyy**