

Analisis Sentimen menggunakan Metode *Naïve Bayes Classifier* terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur *N-gram* dan *Document Frequency Thresholding*

Sinta Kusuma Wardani¹, Yuita Arum Sari², Indriati³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹sintakusuma@student.ub.ac.id, ²yuita@ub.ac.id, ³indriati.tif@ub.ac.id

Abstrak

Pengaruh budaya dan gaya hidup yang semakin berkembang ini membuat masyarakat lebih memperhatikan penampilan. Salah satu faktor yang mempengaruhi penampilan adalah kondisi kulit wajah seseorang. Masing-masing produk yang digunakan konsumen memiliki reaksi yang berbeda-beda antar satu konsumen dengan konsumen lainnya sehingga membuat banyak konsumen melakukan *review* terhadap produk yang digunakan. *Review* yang diberikan oleh konsumen dapat digunakan untuk mengukur kualitas dari sebuah produk kecantikan. Namun banyaknya *review* yang diberikan membuat pengelompokan *review* tidak dapat dilakukan secara manual dan harus dilakukan otomatisasi analisis sentimen untuk mengelompokkan *review* tersebut dalam beberapa kategori. Salah satu algoritme untuk melakukan klasifikasi analisis sentimen, yaitu menggunakan metode *Naïve Bayes Classifier* di mana merupakan metode sederhana yang memiliki performa yang cepat dalam melatih data, mudah dalam implementasinya, serta memiliki efektifitas yang tinggi. Dalam proses klasifikasi akan digunakan seleksi fitur menggunakan algoritme *N-gram* dan *DF-Thresholding* untuk mengurangi dimensi fitur pada data. Tujuan dari penelitian ini untuk mengetahui pengaruh penggunaan algoritme *DF-Thresholding* pada tingkat akurasi algoritme *Naïve Bayes Classifier* menggunakan metode *N-gram*. Hasil penelitian menunjukkan adanya pengurangan fitur sebanyak 16.312 fitur menjadi 43 fitur dengan nilai akurasi tertinggi pada kombinasi *unigram* dan *bigram*, yaitu sebesar 49%, *precision* sebesar 0,23, *recall* sebesar 0,26 serta *f-measure* sebesar 0,24.

Kata kunci: analisis sentimen, *review* produk kecantikan, *Naïve Bayes Classifier*, *N-gram*, *Document Frequency Thresholding*

Abstract

The influence of this growing culture and lifestyle makes people pay more attention to their appearance. One of the factors that affect appearance is the condition of a person's facial skin. Each product used by consumers has different reactions from one consumer to another, thus making many consumers review the products they use. Reviews given by consumers can be used to measure the quality of a beauty product. However, the large number of reviews given makes the review grouping unable to be done manually and sentiment analysis must be done to group the reviews into several categories. One of the algorithms for classifying sentiment analysis is using the *Naïve Bayes Classifier* method which is a simple method that has faster performance in training data, is easy to implement, and has high effectiveness. In the classification process, feature selection will be used using the *N-gram* algorithm and *DF-Thresholding* to reduce the dimensions of the features in the data. The purpose of this study is to determine the effect of *DF-Thresholding* algorithm on the accuracy of the *Naïve Bayes Classifier* algorithm using the *N-gram*. The result showed a reduction of 16.312 features to 43 features and the highest accuracy value for *bigram* and *unigram* combination, which is 49%, *precision* is 0,23, *recall* is 0,26 and *f-measure* is 0,24.

Keywords: sentiment analysis, beauty product reviews, *Naïve Bayes Classifier*, *N-gram*, *Document Frequency Thresholding*

1. PENDAHULUAN

Perubahan gaya hidup dan pengaruh budaya yang semakin berkembang dalam kehidupan sosial membuat masyarakat lebih memperhatikan penampilan. Salah satu faktor yang mempengaruhi penampilan adalah kondisi kulit wajah seseorang. Keadaan kulit wajah yang bersih, sehat dan elok dapat membuat penampilan seseorang lebih menarik dan menawan.

Rasa ingin tahu dari konsumen akan *klaim* pada setiap produk perawatan kulit wajah mendorong konsumen untuk melakukan *review* dari setiap produk yang dikeluarkan oleh suatu *brand*. *Review* atau ulasan yang diberikan merupakan pengalaman pribadi konsumen mengenai kelebihan dan kekurangan produk serta reaksinya pada kulit ketika menggunakan produk tertentu. Seperti halnya pada situs FemaleDaily.com yang merupakan situs yang bergerak dibidang kecantikan, dimana didalamnya terdapat ulasan mengenai beberapa produk yang ada (Rosi, Fauzi dan Perdana, 2018). Pada situs FemaleDaily.com terdapat lebih dari 100.000 ulasan yang diberikan oleh konsumen selaku pemakai produk.

Banyaknya ulasan tentang produk kecantikan dapat digunakan sebagai evaluasi mengenai kualitas produk dan kepuasan konsumen bagi sebuah *brand* atau perusahaan yang mengeluarkan produk tertentu. Ulasan produk tersebut sangat penting bagi produsen sebagai masukan untuk memperbaiki kualitas dari produk mereka kedepannya (Pujadayanti, Fauzi dan Sari, 2018). Disamping itu, ulasan yang diberikan juga dapat membantu konsumen untuk melakukan penilaian sebelum membeli produk yang diinginkan. Hal ini penting untuk dilakukan klasifikasi ulasan dalam kelas positif, negatif maupun netral sehingga dapat diketahui tingkat kualitas dari sebuah produk perawatan wajah berdasarkan nilai dari masing-masing kelas klasifikasi. Jumlah ulasan yang besar menyebabkan proses klasifikasi ulasan tidak dapat dilakukan secara manual. Maka dari itu perlu dilakukan analisis sentimen secara otomatis pada ulasan produk sehingga dapat digunakan sebagai kajian untuk perusahaan terkait dalam melakukan perbaikan dan inovasi baru. Analisis sentimen juga dapat digunakan sebagai bahan pertimbangan konsumen dalam pembelian produk yang cocok dan tepat.

Dalam proses klasifikasi pada analisis sentimen ulasan produk perawatan kulit wajah, metode yang digunakan adalah *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* merupakan metode sederhana yang memiliki performa yang cepat dalam melatih data, mudah dalam implementasinya, serta memiliki efektifitas yang tinggi (Taheri dan Mammadov, 2015). Penelitian penggunaan metode *Naïve Bayes* untuk analisis sentimen sebelumnya pernah dilakukan oleh Antinasari, Perdana dan Fauzi (2017) dimana penggunaan metode *Naïve Bayes Classifier* dengan perbaikan kata tidak baku pada analisis sentimen menghasilkan nilai *accuracy* 98,33%, *precision* 96,77%, *recall* 100% dan *f-measure* 98,36%.

Dalam sebuah penelitian menunjukkan bahwa penggunaan kombinasi kata dalam *N-gram* pada metode *Naïve Bayes* cukup efektif menyelesaikan masalah klasifikasi dokumen. Akurasi terbaik yang didapat sebesar 97% dengan menggunakan teknik *full pre-processing* (Pujadayanti, Fauzi dan Sari, 2018). Namun terdapat permasalahan yang sangat umum terkait dengan penggunaan *N-gram* sebagai seleksi fitur yaitu tingginya dimensi dari fitur yang dihasilkan dalam proses klasifikasi. Tingginya dimensi data yang dihasilkan dapat membuat proses analisis yang dilakukan menjadi lambat dan mengurangi tingkat akurasi dari fitur yang dihasilkan. Sehingga untuk meminimalisir kelemahan dari algoritme *N-gram* maka diperlukan algoritme tambahan untuk mengurangi dimensi data dan *noise* yang tidak diharapkan. Pada penelitian yang dilakukan oleh Rahman, Wiranto dan Doewes (2017) penggunaan algoritme *Document Frequency Thresholding* sebagai seleksi fitur terbukti dapat digunakan untuk melakukan reduksi dimensi dataset yang besar. Selain itu, penggunaan algoritme *DF-Thresholding* dapat digunakan untuk mengurangi komputasi pada perhitungan data.

2. LANDASAN KEPUSTAKAAN

2.1. Review Produk

Review produk berisi mengenai ulasan sebuah produk yang menjelaskan mengenai kualitas dan karakteristik dari produk tertentu berdasarkan pengalaman pribadi dari *customer*. Ulasan yang diberikan oleh *reviewer* mengenai produk tertentu dapat memberikan gambaran kualitas maupun kuantitas dari produk tersebut

sehingga dapat mempengaruhi kepercayaan masyarakat sebagai konsumen (Nabila 2019)

2.2. Analisis Sentimen

Analisis sentimen atau juga disebut *opinion mining* merupakan bidang studi yang melakukan analisis terhadap pendapat, sentimen, evaluasi, penilaian, sikap dan emosi seseorang terhadap entitas tertentu seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan lainnya (Liu 2015).

2.3. Text Mining

Textmining merupakan metode pengolahan teks dengan menggunakan konsep yang ada pada data mining untuk mencari pola dari sebuah dokumen dan melakukan analisis hubungan antara satu dokumen dengan dokumen lainnya.

Dalam penerapannya, text mining melakukan identifikasi dan eksplorasi pola unik untuk mendapatkan informasi penting dari sebuah dokumen (Yudho Baskoro, Ridok dan Tanzil Furqon, 2015).

2.4. Text Preprocessing

Text preprocessing merupakan tahap mengubah bentuk data yang belum terstruktur berupa *term index* yang mewakili dataset dokumen. Menurut Mulyani, Muhamad dan Cahyanto, 2021 tahap preprocessing melalui beberapa tahap yaitu :

1. *Case folding* adalah tahap yang digunakan untuk mengubah huruf kapital pada dokumen menjadi huruf kecil (*lowercase*). Selain itu, pada tahap ini juga dilakukan penghapusan angka, tanda baca dan *whitespace* (karakter kosong).
2. *Tokenizing* adalah tahap pemotongan string berdasarkan kata yang menyusun sebuah dokumen.
3. *Filtering (stopword removal)* adalah tahap mengambil kata-kata penting dari hasil *tokenizing* dan penghilangan kata-kata yang dianggap tidak memiliki makna.
4. *Stemming* adalah proses menghilangkan infleksi kata ke bentuk dasarnya.

2.5. Term Weighting

Pembobotan kata yang digunakan adalah *log TF (Term Frequency)* dimana pembobotan kata *log TF* akan menghitung berdasarkan kemunculan *term t* pada dokumen *d* (Nurjanah, Perdana dan Fauzi, 2017). Rumus *log TF* pada persamaan (1)

$$w_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d} & \text{if } tf_{t,d} > 0 \\ 0 & \text{if } tf_{t,d} = 0 \end{cases} \quad (1)$$

Keterangan :

$w_{t,d}$ = *term frequency* dengan pendekatan *term frequency*

$tf_{t,d}$ = jumlah kemunculan *term t* pada dokumen *d*

2.6. Metode N- gram

N-gram merupakan potongan n-karakter yang diperoleh dari sebuah string yang diaplikasikan guna untuk pembangkitan kata atau karakter (Prasanti, Fauzi dan Furqon, 2018).

Contoh pada kata “TEKS”.

Unigram : T,E,K,S

Bigram : TE, EK,KS

Trigram : TEK dan EKS

2.7. Document Frequency Thresholding

Document Frequency Thresholding merupakan metode seleksi fitur *unsupervised* yang memiliki bentuk sederhana dan komputasi yang rendah. Metode ini memiliki asumsi bahwa term yang lebih jarang muncul tidak memiliki pengaruh besar dalam proses pengelompokan dokumen, sehingga dapat mengurangi dimensi fitur yang besar dalam text mining (Langgeni, Baizal dan Firdaus, 2010).

2.8. Naïve Bayes Classifier

Naïve Bayes Classifier merupakan teknik klasifikasi yang dapat bekerja dengan cepat dengan akurasi yang tinggi pada jumlah data yang besar. Dimana asumsinya bahwa nilai *attribut* pada kelas tertentu tidak tergantung pada nilai *attribut* lainnya (Pujadayanti, Fauzi dan Sari, 2018). Rumus dapat dilihat pada persamaan (3).

$$P(C_j|W_i) = P(C_j) \prod_{1 \leq i \leq nd} P(w_i|C_j) \quad (3)$$

Dimana :

$P(C_j|W_i)$: *posterior* merupakan menghitung peluang kemunculan kelas C_j dimana $j = 1,2,3 \dots$

$P(w_i|C_j)$: *conditional probability* adalah menghitung peluang kemunculan *term* kata pada kelas j .

$P(C_j)$: *prior* adalah menghitung peluang kemunculan dokumen kelas j pada

data latih.

2.9. Multinomial Naïve Bayes

Multinomial naïve bayes merupakan perhitungan bersifat *positional independence* tidak tergantung pada posisi ataupun urutan kata. Proses keseluruhan pada metode ini tidak jauh berbeda dengan metode *Naïve Bayes Classifier* yaitu perhitungan *prior*, *conditional probability* dan *posterior* akan tetapi untuk *conditional probability* terdapat perubahan (Manning, Raghavan dan Schutze, 2009). Rumus dapat dilihat pada persamaan (4)

$$P(C_j|W_i) = \frac{\text{Count}(w_i|C_j)}{\sum_{w \in V} \text{Count}(w, C_j) + |V|} \quad (4)$$

Dimana :

$P(C_j|W_i)$: *conditional probability*

$\text{Count}(w_i|C_j)$: menghitung jumlah kata W_i dengan $i = 1, 2, 3, \dots, m$ pada kelas j dengan $j = 1, 2, 3, \dots, n$

$\text{Count}(W, C_j)$: menghitung jumlah seluruh kata kelas j

$|V|$: menghitung kata unik pada seluruh dokumen.

2.10. Evaluasi

proses evaluasi digunakan untuk mengetahui nilai *accuracy*, *precision*, *recall*, dan *f-measure* yang dinyatakan pada tabel *confusion matrix multiclass* yang dapat dilihat pada Tabel 1.

Tabel 1 Confusion Matrix

Confusion Matrix		Kenyataan		
		Positif	Negatif	Netral
Prediksi	Positif	TP	E ₁₂	E ₁₃
	Negatif	E ₂₁	TP	E ₂₃
	Netral	E ₃₁	E ₃₂	TP

Dengan rumus perhitungan sebagai berikut :

$$\text{Accuracy (AC)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$\text{Precision (P)} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall (R)} = \frac{TP}{TP+FN} \quad (7)$$

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

2.11. Evaluasi Relevansi

Evaluasi relevansi merupakan evaluasi yang dilakukan untuk menghitung nilai relevansi dari dokumen yang dijadikan *dataset*. *Kappa measure* digunakan untuk mengukur kesepakatan antara para penilai pada sebuah dokumen (Han, Kamber dan Pei, 2012). Nilai *kappa measure* dapat dihitung menggunakan

persamaan (9).

$$Kappa = \frac{P(A)-P(E)}{1-P(E)} \quad (9)$$

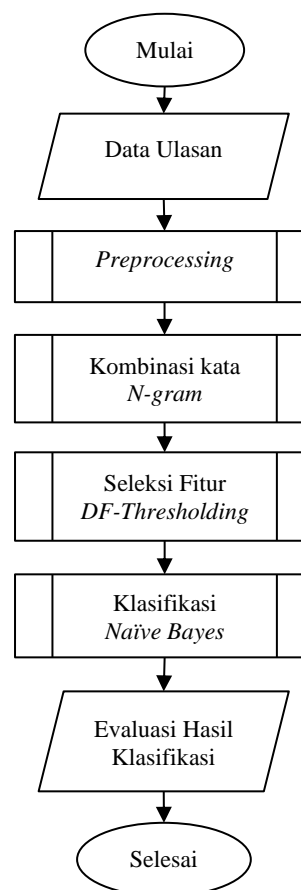
Keterangan :

$P(A)$: Proporsi banyaknya penilai yang sama

$P(E)$: Proporsi banyaknya penilai yang sama karena kebetulan

3. METODE

Berdasarkan Gambar 1 menjelaskan tahapan-tahapan analisis sentimen yang dilakukan pada penelitian ini. Pada tahap *preprocessing* akan dilakukan sub proses lagi seperti yang telah dijelaskan sebelumnya. Begitupun dengan proses kombinasi kata *N-gram*, seleksi fitur *DF-Thresholding*, dan klasifikasi *Naïve Bayes*.



Gambar 1 Diagram Alir Sistem

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data ulasan produk beberapa produk skincare dan brand berbeda yang diambil dari situs web www.FemaleDaily.com. Data yang diambil terdiri atas 3 kategori produk yakni

kategori pelembab wajah, sabun muka, dan *toner*. Dari masing-masing kategori yang diambil terdiri 5 produk dari *brand* yang berbeda. Data yang digunakan berbentuk tabel dengan fitur sebanyak 7 fitur dan jumlah 1068 data ulasan.

3.2 Perancangan Algoritme

Pada tahap pertama data akan diberikan label positif, negatif dan netral. Kemudian data data ulasan akan mengalami *preprocessing* dengan tujuan merubah data ulasan menjadi *term index*. Preprocessing yang dilakukan meliputi *case folding*, *stemming*, *tokenizing*, normalisasi bahasa, dan *filtering*. Sebelum dilakukan tahap *filtering*, akan dilakukan tahap normalisasi bahasa dengan kamus bahasa *alay* (Ibrohim dan Budi, 2019). Tujuan tahapan normalisasi bahasa ini adalah untuk mengubah kata tidak baku menjadi kata baku sesuai KBBI. Kemudian setelah itu akan dilakukan proses *filtering* (*stopword removal*).

Data yang telah dilakukan *preprocessing* akan berupa *term* yang kemudian akan dilakukan kombinasi kata menggunakan algoritme *N-gram* dan seleksi fitur dengan algoritme *DF-Thresholding* untuk mengurangi banyaknya dimensi data. Pembobotan kata yang digunakan adalah pembobotan *log TF* yang kemudian dokumen akan diklasifikasikan kedalam sentimen positif, negatif, dan netral menggunakan algoritme *Naïve Bayes Classifier*.

Pada proses klasifikasi data yang digunakan akan dibagi sebesar 70% untuk data *training* dan 30% untuk data *testing*. Data *training* yang telah melewati proses

preprocessing dan seleksi fitur oleh algoritme *N-gram* dan *DF-Thresholding* akan digunakan sebagai bahan pembelajaran dalam menentukan kelas klasifikasi pada data *testing*.

Data *testing* yang digunakan akan dilakukan tahap *preprocessing* terlebih dahulu sebelum dilakukan klasifikasi. Tahap-tahap klasifikasi menggunakan *Naïve Bayes* adalah sebagai berikut :

1. Menghitung nilai *prior* masing-masing kelas positif, negatif, dan netral
2. Melakukan pembobotan kata menggunakan algoritme *log TF*
3. Menghitung nilai *likelihood* setiap kelas menggunakan rumus *Multinomial Naïve Bayes*
4. Menghitung nilai *posterior* dari dokumen *testing* untuk menentukan kelas klasifikasi
5. Menentukan kelas dokumen ulasan berdasarkan nilai tertinggi dari *posterior*

4. HASIL DAN PENGUJIAN

4.1 Document Frequency Thresholding

Pada tahap *DF-Thresholding* akan dilakukan pengujian terhadap batasan *threshold* yang digunakan yaitu nilai *minDF*, *maxDF*, *minFreq*, dan *maxFreq*.

Berdasarkan pengujian seleksi fitur menggunakan algoritme *DF-Thresholding* yang dilakukan pada data *training* maka diperoleh hasil seperti yang ditunjukkan pada Tabel 2.

Tabel 2 Pengujian *DF-Thresholding*

N o	minDF	maxDF	minFreq	maxFreq	Precision	Recall	F- measure	Accuracy (%)
1	0	0,8	0,25	0,7	0,23	0,26	0,24	49%
2	0,2	0,8	0,2	0,8	0,2	0,23	0,21	41,81%
3	0,2	0,8	0,3	0,7	0,2	0,26	0,22	43,44%
4	0,2	0,8	0,25	0,7	0,2	0,25	0,22	44,37%
5	0,25	0,75	0,35	0,7	0,2	0,28	0,22	42,81%
6	0,35	0,7	0,25	0,75	0,2	0,28	0,22	41,56%
7	0,4	0,6	0,25	0,7	0,19	0,27	0,21	38,12%

Berdasarkan Tabel 2 dapat diketahui bahwa nilai *threshold* terbaik adalah pada rentang 0 dan

0,8 untuk dokumen frekuensi, serta rentang 0,25 dan 0,7 untuk *term* frekuensi, dengan nilai

accuracy sebesar 49%, *precision* sebesar 0,23, *recall* sebesar 0,26 dan *f-measure* sebesar 0,24. Rentang nilai *threshold* dokumen frekuensi yang lebih tinggi daripada rentang nilai *threshold term* frekuensi menghasilkan akurasi lebih tinggi daripada pengujian yang memiliki rentang nilai *threshold* dokumen frekuensi yang lebih rendah. Hal ini disebabkan semakin tinggi nilai *threshold* yang ditetapkan, semakin sedikit kata yang lolos sehingga memungkinkan banyak kata penting yang terbuang.

4.2 Evaluasi

Pada proses evaluasi akan dilakukan pengujian antara dua kondisi, yaitu proses klasifikasi *Naïve Bayes* menggunakan *N-gram* dan *DF-Thresholding* dengan klasifikasi *Naïve Bayes* menggunakan *N-gram* tanpa *DF-Thresholding*. Pengujian ini dilakukan untuk mengetahui efektifitas metode seleksi fitur yang digunakan. Hasil evaluasi menggunakan seleksi fitur *DF-Thresholding* ditunjukkan pada Tabel 3

Tabel 3 klasifikasi Penggunaan *N-gram* dan *DF-Thresholding*

Uji	Menggunakan <i>DF-Thresholding</i>			
	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Accuracy (%)</i>
<i>unigram</i>	0,2	0,25	0,22	44,37%
<i>bigram</i>	0,17	0,26	0,2	36,25%
Kombinasi <i>unigram</i> dan <i>bigram</i>	0,23	0,26	0,24	49%

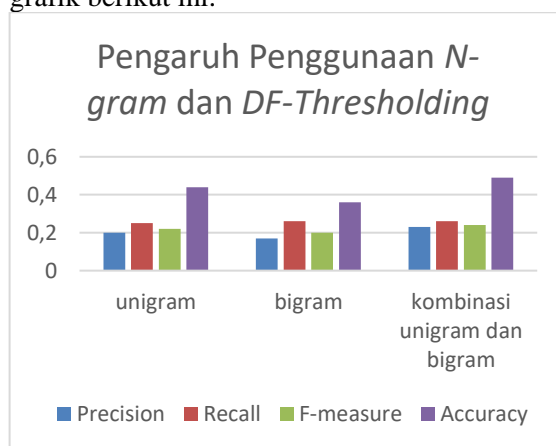
Hasil evaluasi tanpa seleksi fitur *DF-Thresholding* ditunjukkan pada Tabel 4.

Tabel 4 klasifikasi Penggunaan *N-gram* tanpa *DF-Thresholding*

Uji	Tanpa <i>DF-Thresholding</i>			
	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Accuracy (%)</i>
<i>unigram</i>	0,26	0,3	0,27	55%
<i>bigram</i>	0,24	0,32	0,26	50,94%

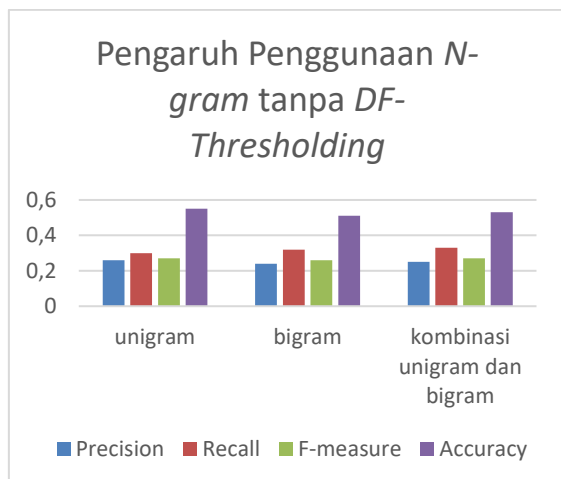
Kombinasi <i>unigram</i> dan <i>bigram</i>	0,25	0,33	0,27	53,12%
--	------	------	------	--------

Untuk mengetahui perbandingan antara 2 pengujian diatas, dapat dilihat pada gambar grafik berikut ini.



Gambar 2 Penggunaan *N-gram* dan *DF-Thresholding*

Pada Gambar 2 grafik pengujian kombinasi *unigram* dengan *bigram* memiliki grafik yang lebih tinggi dari pengujian menggunakan *unigram* atau *bigram*. Pada pengujian *bigram* memiliki nilai *accuracy*, *precision*, *recall* dan *f-measure* yang lebih rendah dikarenakan sedikitnya kombinasi kata *bigram* pada *term data training* dengan *term data testing* yang sesuai. Pada pengujian penggunaan algoritme *N-gram* dengan algoritme seleksi fitur *DF-Thresholding* menghasilkan nilai *accuracy* tertinggi sebesar 49%, *precision* sebesar 0,23, *recall* sebesar 0,26 dan *f-measure* sebesar 0,24 pada kombinasi *unigram* dan *bigram*. Hal ini disebabkan karena banyaknya *term* yang cocok dengan *term data training* kombinasi *unigram* dan *bigram* sehingga banyak data yang tepat klasifikasi.



Gambar 3 Penggunaan N-gram tanpa DF-Thresholding

Gambar 3 menunjukkan hasil pengujian pengaruh penggunaan *N-gram* tanpa seleksi fitur menggunakan metode *DF-Thresholding*. Hasil tertinggi adalah menggunakan kata *unigram* dengan nilai *accuracy* sebesar 55%, *precision* sebesar 0,26, *recall* sebesar 0,3 serta *f-measure* sebesar 0,27. Dibandingkan dengan pengujian sebelumnya, pengujian tanpa menggunakan *DF-Thresholding* menghasilkan nilai *accuracy* yang lebih tinggi dibanding dengan penggunaan *DF-Thresholding* sebagai seleksi fitur. Terdapat beberapa *term* yang mewakili dokumen tertentu hilang ketika terjadi pengurangan *term* berdasarkan nilai *threshold* yang telah ditetapkan menyebabkan akurasi yang dihasilkan justru menurun. Disamping hasilnya yang kurang dalam meningkatkan performa klasifikasi, algoritme *DF-Thresholding* termasuk efektif dalam mengurangi dimensi data dengan mengurangi *term* yang memiliki nilai frekuensi rendah sehingga dapat mengurangi proses komputasi.

4.3 Evaluasi Relevansi

Pengujian ini digunakan untuk menghitung kesepakatan antar penilai menggunakan *kappa-measure*. Pada penelitian ini menggunakan sebanyak tiga penilai sebagai perbandingan yang kemudian akan dihitung berapa nilai kesepakatan antar penilai. Hasil perhitungan nilai *kappa-measure* antara 3 penilai dapat dilihat pada tabel dibawah ini :

Tabel 5 Hasil Perbandingan Penilai 1 dan 2

	Penilai 2			
	Positif	Negatif	Netral	Total

Penilai 1	Positif	402	6	20	428
	Negatif	0	245	62	307
	Netral	21	16	296	333
	Total	423	267	378	1068
<i>P(A)</i>		0,882959			
<i>P(E)</i>		0,340942			
<i>Kappa-Measure</i>		0,822411			

Tabel 6 Hasil Perbandingan Penilai 1 dan 3

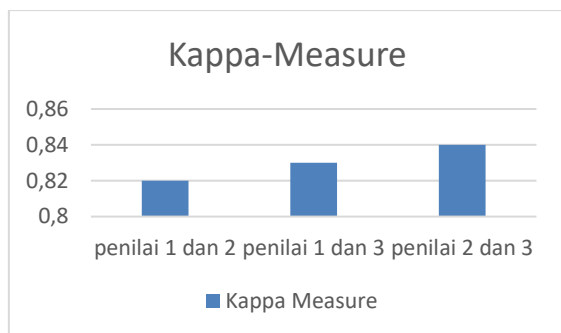
		Penilai 3			
		Positif	Negatif	Netral	Total
Penilai 1	Positif	400	5	23	428
	Negatif	1	256	50	307
	Netral	22	16	295	333
	Total	423	277	368	1068
<i>P(A)</i>		0,890449			
<i>P(E)</i>		0,340714			
<i>Kappa-Measure</i>		0,833834			

Tabel 7 Hasil Perbandingan Penilai 2 dan 3

		Penilai 3			
		Positif	Negatif	Netral	Total
Penilai 2	Positif	404	0	19	423
	Negatif	2	236	29	267
	Netral	17	41	320	378
	Total	423	277	368	1068
<i>P(A)</i>		0,898876			
<i>P(E)</i>		0,343665			
<i>Kappa-Measure</i>		0,845927			

Untuk lebih memahami perbandingan pengujian

kappa-measure antara 3 penguji, dapat dilihat pada gambar grafik berikut ini.



Gambar 4 Hasil Perbandingan Nilai Kappa-Measure antar 3 Penilai

Berdasarkan Gambar 4, dapat dilihat bahwa perbandingan yang dilakukan antara penilai 1 dan penilai 2, penilai 1 dan penilai 3 serta penilai 2 dan penilai 3 menghasilkan nilai *kappa-measure* berturut-turut sebesar 0,82, 0,83, dan 0,84. Nilai *kappa-measure* yang dihasilkan oleh ketiga perbandingan tersebut dapat digolongkan dalam kategori baik dengan kriteria nilai diatas 0,8. Hal ini dapat disimpulkan bahwa dokumen-dokumen yang digunakan merupakan *dataset* yang relevan.

5. KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan yang dilakukan, diperoleh kesimpulan bahwa metode *Naïve Bayes Classifier* dapat digunakan untuk melakukan analisis sentimen terhadap ulasan produk perawatan wajah. Hal ini dapat ditunjukkan dengan hasil akhir akurasi yang didapat yaitu sebesar 55%. Tahapan analisis sentimen yang dilakukan meliputi *preprocessing*, kombinasi kata *N-gram*, seleksi fitur menggunakan algoritme *DF-Thresholding*, *term weighting*, dan klasifikasi menggunakan metode *Naïve Bayes Classifier*.

Penggunaan variasi nilai *threshold* terbaik dalam penggunaan seleksi fitur *DF-Thresholding* berada pada rentang nilai *threshold* dokumen frekuensi sebesar 0 sampai 0,8 dan *term* frekuensi sebesar 0,25 sampai 0,7. Pada rentang nilai *threshold* tersebut menghasilkan nilai akurasi tertinggi sebesar 49%, *precision* sebesar 0,23, *recall* sebesar 0,26 dan *f-measure* sebesar 0,24. Rentang nilai *threshold* dokumen frekuensi yang lebih tinggi daripada rentang nilai *threshold term* frekuensi menghasilkan akurasi lebih tinggi daripada pengujian yang memiliki rentang nilai *threshold* dokumen frekuensi yang

lebih rendah.

Penggunaan algoritme *N-gram* pada proses klasifikasi menghasilkan nilai akurasi tertinggi pada kombinasi *unigram* dan *bigram* yaitu sebesar 49%, *precision* sebesar 0,23, *recall* sebesar 0,26 serta *f-measure* sebesar 0,24 dengan menggunakan algoritme *DF-Thresholding*. Sementara penggunaan *N-gram* tanpa seleksi fitur *DF-Thresholding* menghasilkan nilai akurasi tertinggi pada *unigram* yaitu sebesar 55%, *precision* sebesar 0,26, *recall* sebesar 0,3 serta *f-measure* sebesar 0,27. Penggunaan algoritme *N-gram* dengan *DF-Thresholding* menghasilkan akurasi akhir lebih rendah dibandingkan dengan penggunaan algoritme *N-gram* tanpa *DF-Thresholding*. Dari pengujian yang dilakukan menunjukkan bahwa *DF-Thresholding* kurang dapat meningkatkan performa klasifikasi oleh algoritme *Naïve Bayes Classifier*, namun dapat mengurangi dimensi data pada proses klasifikasi yang ditunjukkan dengan pengurangan fitur sebanyak 16.312 fitur menjadi 43 fitur.

6. SARAN

Adapun saran untuk penelitian analisis sentimen produk perawatan wajah menggunakan metode *Naïve Bayes Classifier* serta metode seleksi fitur *N-gram* dan *DF-Thresholding* selanjutnya adalah penggunaan seleksi fitur yang sesuai sehingga kombinasi metode yang digunakan sehingga dapat meningkatkan keakuratan dari penelitian kedepannya. Selain itu, penggunaan metode *term weighting* yang sesuai dan perbaikan kata tidak baku serta bahasa asing yang banyak terdapat pada dokumen ulasan produk.

7. DAFTAR PUSTAKA

- Antinasari, Prananda, Rizal Setya Perdana and M Ali Fauzi. 2017. "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes Dengan Perbaikan Kata Tidak Baku" 1: 1733–41.
- Han, Jiawei, Micheline Kamber and Jian Pei. 2012. *Data Mining: Concepts and Techniques: Concepts and Techniques (3rd Edition)*. Data Mining.
- Ibrohim, Muhammad Okky. 2019. "Multi-Label Hate Speech and Abusive Language Detection in Indonesian Twitter", 46–57.

- Langgeni, Baizal, Firdaus. 2010. "Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection" 2010: 1–10. <<https://doi.org/10.31219/osf.io/tm8c5>>.
- Liu, Bing. 2015. "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions". *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, no. May: 1–367. <<https://doi.org/10.1017/CBO9781139084789>>.
- Mammadov, Musa. 2015. "Learning the Naive Bayes Classifier with Optimization Models", no. December 2013. <<https://doi.org/10.2478/amcs-2013-0059>>.
- Manning, christopher d., Prabhakar Raghavan and Hinrich Schutze. 2009. "An Introduction To Informational Retrieval". *Information Retrieval*.
- Mulyani, Esti, Fachrul Pralienka, Bani Muhamad and Kurnia Adi Cahyanto. 2021. "Pengaruh N-Gram Terhadap Klasifikasi Buku Menggunakan Ekstraksi Dan Seleksi Fitur Pada Multinomial Naïve Bayes" 5: 264–72. <<https://doi.org/10.30865/mib.v5i1.2672>>.
- Nabila, Sheila Mariah. 2019. "PENGARUH REVIEW PRODUK DAN BRAND AWARENESS TERHADAP TRUST DAN MINAT BELI: STUDI KASUS REVIEW FLOW FUSHI ION DE CUSHION OLEH ALLYSSA HAWADI MELALUI INSTAGRAM". *Jurnal Ilmu Komunikasi* 3: 107–21.
- Nurjanah, Winda Estu, Rizal Setya Perdana and Mochammad Ali Fauzi. 2017. "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor Dan Pembobotan Jumlah Retweet". *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya* 1 No. 12: 1750–57.
- Prasanti, Annisya Aprilia, M. Ali Fauzi and M. Tanzil Furqon. 2018. "Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode N- Gram Dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)". *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya* Vol. 2: 594–601.
- Pujadayanti, Irma, Moch Ali Fauzi and Yuita Arum Sari. 2018. "Prediksi Rating Otomatis Pada Ulasan Produk Kecantikan Dengan Metode Naïve Bayes Dan N-Gram" 2: 4421–27.
- Rahman, Amelia, Wiranto Wiranto and Afrizal Doewes. 2017. "Online News Classification Using Multinomial Naive Bayes". *ITSMART: Jurnal Teknologi Dan Informasi* 6: 32–38.
- Rosi, Fathor, M Ali Fauzi and Rizal Setya Perdana. 2018. "Prediksi Rating Pada Review Produk Kecantikan Menggunakan Metode Naïve Bayes Dan Categorical Proportional Difference (CPD)" 2: 1991–97.
- Yudho Baskoro, Setyoko, Achmad Ridok and Muhammad Tanzil Furqon. 2015. "Pencarian Pasal Pada Kitab Undang-Undang Hukum Pidana (Kuhp) Berdasarkan Kasus Menggunakan Metode Cosine Similarity Dan Latent Semantic Indexing (Lsi)". *Journal of Enviromental Engineering and Sustainable Technology* 2: 83–88. <<https://doi.org/10.21776/ub.jeest.2015.02.02.4>>.