

## Rate Prediction of Cosmetic Product Based on Test Review from Website Female Daily Using Naive Bayes Classifier

**Dian Cahya Oktaviana**

Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36A, Surakarta  
dianco@student.uns.ac.id

**Bambang Harjito**

Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36A, Surakarta  
bambang\_harjito@staff.uns.ac.id

**Sari Widya Sihwi**

Informatika, Fakultas MIPA  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36A, Surakarta  
sari.widya.sihwi@gmail.com

### ABSTRAK

Dalam penelitian ini dilakukan prediksi rating terhadap teks review produk kosmetik yang terdapat pada website Female Daily. Penelitian prediksi rating ini menggunakan pendekatan *machine learning*. Algoritma *Naive Bayes Classifier* dipilih karena akurasi dan kesederhanaan dalam penggunaannya.

Klasifikasi rating yang dihasilkan dari penelitian ini mampu mengelompokkan sebuah teks review termasuk ke dalam kelas rating satu, dua, tiga, empat, atau lima. Total dataset yang digunakan sebanyak 688 data yang terdiri dari 550 data *training* dan 138 data *testing*. Eksperimen yang telah dilakukan menunjukkan bahwa metode yang diusulkan memiliki akurasi sebesar 49,27% dengan class precision 49,61 % dan class recall 48,19%. Sementara akurasi rata-rata sebesar 38,62%.

Kata Kunci: *naive bayes classifier*, prediksi rating, text review

### 1. PENDAHULUAN

Saat ini, jika seseorang ingin membeli suatu produk, mereka tidak lagi terbatas hanya bisa bertanya pada teman dan keluarga saja karena ada banyak *online review* dari pengguna produk di dalam forum online [1]. Hal ini terjadi karena perkembangan berbagai macam portal web dan media komunikasi seperti situs jejaring sosial, forum, dan blog yang membuat orang semakin mudah untuk saling bertukar pikiran dan berdiskusi mengenai berbagai macam topik bahasan.

*Female Daily* merupakan sebuah forum dan blog Indonesia dengan konten khusus tentang kecantikan. Tercatat melalui laman <http://reviews.femaledaily.com/>, portal tersebut memiliki lebih dari 10.000 *review* tentang produk kecantikan dan dihuni oleh lebih dari 230 ribu anggota [2]. Website ini terdiri dari 3 kolom yaitu *review*, blog dan forum. Pada kolom *review* pengguna terdiri dari dua bagian, yaitu nilai rating dan komentar *review* secara tekstual tentang produk yang di-*review*.

*Review* atau ulasan Female Daily adalah sumber informasi yang penting bagi pengguna produk kosmetik sebelum menentukan produk kosmetik yang akan dibeli diantara berbagai macam pilihan yang tersedia. Namun

karena banyaknya ulasan, hampir tidak mungkin bagi pengguna untuk membaca semua ulasan dan menemukan informasi yang mereka cari. Untuk itu dalam penelitian ini akan dilakukan prediksi rating.

*Rating Prediction* atau populer dengan sebutan *sentiment rating prediction* adalah cabang penelitian yang berkaitan dengan pemberian nilai rating pada data tekstual [3]. Tugas utamanya yaitu bertujuan untuk memprediksi skor *review* berdasarkan sebuah teks *review*.

Prediksi Rating atau *Rating Prediction* dianggap sebagai permasalahan *multi-class classification* dimana kelas menjadi skor atau nilai rating dari produk [4], sehingga masalahnya dapat diselesaikan dengan menggunakan metode klasifikasi dokumen *machine learning* seperti *Naive Bayes Classifier* (NBC), *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), dan lain-lain [5].

Penelitian yang berkaitan dengan masalah pengklasifikasian dokumen teks sebelumnya telah banyak dilakukan, diantaranya yaitu dilakukan oleh [6] yang mencoba menyelesaikan permasalahan analisis sentimen dengan teknik klasifikasi, yaitu membandingkan antara algoritma *Support Vector Machine* dan *Naive Bayes* untuk mengklasifikasikan *review* restoran dalam bahasa Cantonese. Dalam penelitian tersebut metode *Naive Bayes* menunjukkan performa lebih baik. Dalam penelitian lain, [2] melakukan analisis sentimen data twitter menggunakan metode *Naive Bayes Classifier* dan seleksi fitur *N-gram*.

Penelitian yang berkaitan dengan prediksi rating juga telah dilakukan sebelumnya, diantaranya pernah dilakukan oleh [7] dengan mencoba menggunakan metode *Linear Regression*, *Support Vector Regression*, dan *Decision Tree Regression*. Pada tahun 2014, [8] menerapkan metode *supervised learning* untuk memprediksikan rating dari *review* dengan cara memberikan skala numerik berdasarkan data teks. Metode yang digunakan dalam penelitian tersebut antara lain adalah *Naive Bayes*, *Perceptron*, *Nearest Neighbor* dan *Multiclass Support Vector Machine*. Dalam penelitian itu metode *Naive Bayes* menunjukkan performa terbaik diantara metode-metode yang lainnya. Atas dasar hal ini, peneliti bermaksud menerapkan metode *Naive Bayes Classifier*

untuk memprediksi rating produk kosmetik berdasarkan teks review.

## 2. TEXT PREPROCESSING

Menurut [9] dalam bukunya, *Text Mining* adalah sebuah area penelitian variasi dari teknik *data mining* yang bertujuan untuk menemukan informasi atau pola tertentu dari data tekstual yang berjumlah besar.

Salah satu kegiatan penelitian dalam text mining adalah preprocessing pada data teks, yang bertujuan untuk menyederhanakan dimensi data atau menghilangkan *noise* agar dataset siap digunakan untuk proses selanjutnya, yaitu klasifikasi. Selain itu, proses ini juga dilakukan untuk mendapatkan akurasi yang lebih tinggi dari hasil klasifikasi [10]. *Text preprocessing* terpusat pada identifikasi dan ekstraksi fitur. *Text preprocessing* mentransformasikan data tidak terstruktur yang tersimpan di dalam dokumen menjadi data dengan format yang lebih terstruktur [9]. *Text preprocessing* akan merubah teks menjadi sebuah set *term index* yang bisa mewakili dokumen.

Proses text preprocessing dalam penelitian ini ada 4 tahap, yaitu *case folding*, *filtering*, *stemming* dan *tokenization*.

## 3. PREDIKSI RATING

*Rating Prediction* atau populer dengan sebutan *sentiment rating prediction* adalah cabang penelitian yang berkaitan dengan pemberian nilai rating pada data tekstual [3].

Tugas utamanya yaitu bertujuan untuk memprediksi skor review dari sebuah teks review. Tugas prediksi rating dapat juga dianggap sebagai permasalahan klasifikasi karena skor review hasil prediksi merupakan nilai rating yang dijadikan sebagai kelas klasifikasi.

## 4. NAÏVE BAYES CLASSIFIER

*Naive Bayes Classifier* merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan Teorema *Bayes* dengan asumsi ketidaktergantungan (*independen*) yang tinggi. Penggunaan metode *Naive Bayes Classifier* pada penelitian ini didasarkan pada banyaknya dataset yang dipakai sehingga membutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi [11]. Keuntungan penggunaan *Naive Bayes Classifier* adalah metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Metode *Naive Bayes Classifier* menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses terhadap sampel data yang sedapat mungkin dapat menjadi representasi data tersebut. Selanjutnya adalah penentuan probabilitas *prior* bagi tiap kategori berdasarkan sampel

data. Pada tahap klasifikasi ditentukan nilai kategori dari suatu data berdasarkan *term* yang muncul dalam data yang diklasifikasi. Teorema *Naive Bayes* dapat dinyatakan dalam persamaan 1.

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \quad (1)$$

Dimana, keadaan *Posterior* (Probabilitas  $X_k$  di dalam  $Y$ ) dapat dihitung dari keadaan *prior* (Probabilitas  $Y$  di dalam  $X_k$  dibagi dengan jumlah dari semua probabilitas  $Y$  di dalam semua  $X_i$ ).

Untuk dapat mengklasifikasikan suatu *review*, dalam penelitian ini penulis menggunakan metode *Naive Bayes Classifier* untuk klasifikasi teks, seperti yang dilakukan [12] pada persamaan 2.

$$P(v1|C = c) = \frac{\text{CountTerms}(v1, \text{docsv}(c))}{\text{AllTerms}(\text{docs}(c))} \quad (2)$$

Dimana  $v1$  dalam penelitian ini adalah satu kata tertentu dalam *review*, sedangkan  $\text{CountTerms}(v1, \text{docsv}(c))$  menunjuk pada jumlah kemunculan suatu kata berlabel  $c$  ("satu" atau "dua" atau "tiga" atau "empat" atau "lima").  $\text{AllTerms}(\text{docs}(c))$  menunjuk pada jumlah semua kata berlabel  $c$  yang ada pada dataset. Untuk menghindari adanya nilai nol pada probabilitas, maka diberlakukan *Laplace (add-one) smoothing*. Tujuan daripada *smoothing* adalah untuk mengurangi probabilitas 0 dari hasil/keluaran yang terobeservasi, dan juga sekaligus meningkatkan/menambah probabilitas hasil/keluaran yang belum terobservasi [13], sehingga persamaan menjadi sebagai berikut:

$$P(v1|C = c) = \frac{\text{CountTerms}(v1, \text{docsv}(c)) + 1}{\text{AllTerms}(\text{docs}(c)) + |V|} \quad (3)$$

Dimana  $|V|$  menunjuk pada jumlah semua kata dalam *review* yang ada di dataset.

## 5. CONFUSION MATRIX

*Confusion matrix* digunakan untuk menampilkan informasi dari hasil klasifikasi berupa *actual class* dan *predicted class*. Performa dari sistem klasifikasi secara umum dievaluasi dengan metrik data [14]. Contoh *confusion matrix* dengan dua kelas ditunjukkan pada tabel 1.

Tabel 1 Contoh Tabel Confusion Matrix Dua Kelas

	Predicted	C1	C2
Actual Class			
C1		TP	FN
C2		FP	TN

Pengukuran evaluasi kinerja yang sering digunakan adalah *accuracy*. *Accuracy* tidak membedakan antara jumlah label

yang benar dari kelas yang berbeda. *Accuracy* dapat dihitung dengan persamaan (4) [15].

$$accuracy = \frac{tp+tn}{tp+fp+fn+tn} \quad (4)$$

Pengukuran kinerja lain yang dipilih untuk kelas positif adalah *precision*, *recall*, dan *error rate*.

*Precision* dihitung dengan persamaan (5) [15]

$$precision = \frac{tp}{tp+fp} \quad (5)$$

*Precision* adalah sebuah fungsi dari contoh positif yang diklasifikasikan dengan benar (*true positive*) dan contoh bernilai negatif salah diklasifikasikan sebagai positif (*false positive*).

$$recall = \frac{tp}{tp+fn} \quad (6)$$

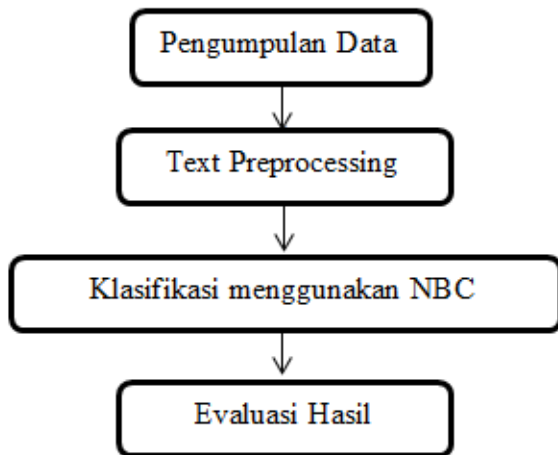
*Recall* adalah sebuah fungsi dari contoh positif yang diklasifikasikan dengan benar (*true positive*) dan contoh positif yang salah diklasifikasikan sebagai negatif (*false negative*) [15].

$$Error\ Rate = \frac{FP+FN}{P+N} \times 100\% \quad (7)$$

*Error rate* didefinisikan sebagai jumlah kesalahan dalam klasifikasi data.

## 6. METODOLOGI

Tahapan dari penelitian ini tersaji dalam Gambar 1.



Gambar 1 Metodologi Penelitian

### 6.1 Pengumpulan Data

Tahap awal dari penelitian ini adalah pengumpulan data. Data ini nantinya akan dibagi kedalam dua bagian, yaitu data untuk *training* dan *testing*. Data didapatkan dari halaman <http://reviews.femaledaily.com> yang berisi kumpulan review pengguna produk kosmetik. Data produk kosmetik yang akan digunakan adalah review produk dari kategori *Shampoo*. Jumlah produk yang digunakan sebanyak 5 produk yang berasal dari *brand* berbeda. Data yang diambil adalah data review berbahasa Indonesia. Akan tetapi ada beberapa data yang mengandung kosakata

bahasa Inggris. Selanjutnya data penelitian dimasukkan kedalam database MySQL. Data ini kemudian akan diolah pada tahap *text processing* untuk menjadikan data agar lebih terstruktur.

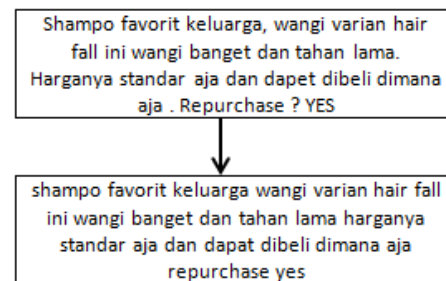
### 6.2 Text Preprocessing

Setelah mendapatkan data, proses selanjutnya yaitu tahap *preprocessing* agar data lebih terstruktur dengan cara menghilangkan *noise*. Proses *preprocessing* dalam penelitian ini dibagi menjadi tiga bagian.

Berikut adalah tahapan preprocessing dalam penelitian ini:

#### a. Case Folding

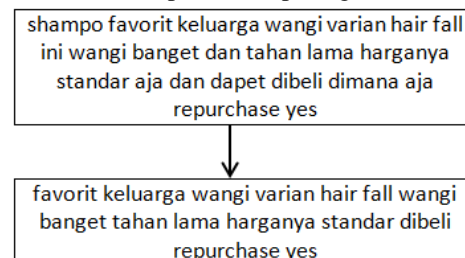
Proses *case folding* atau penghapusan karakter selain huruf yang dianggap delimiter dan mengubah semua karakter menjadi huruf kecil / *lowercase*. Gambar 2 menunjukkan salah satu contoh tahap *case folding* pada teks review.



Gambar 2 Contoh Case Folding

#### b. Filtering / Stopword Removal

Tahap *filtering* adalah penghapusan kata atau *term* yang sering muncul dan tidak memiliki makna inti dari sebuah kalimat yang disebut dengan *stopword*. *Stopword* adalah kosakata yang jumlah kemunculannya besar namun tidak memiliki makna inti terhadap kalimat atau frasa [17]. *Stopword list* yang digunakan dalam penelitian ini terdiri dari kosakata bahasa Indonesia yang berasal dari penelitian yang telah dilakukan oleh [18]. Dan karena dalam dataset review terdapat data yang mengandung kalimat berbahasa Inggris, maka penulis menambahkan beberapa kosakata bahasa Inggris kedalam *stopword list*. Contoh proses *filtering* atau *stopword removal* dapat dilihat pada gambar 3.

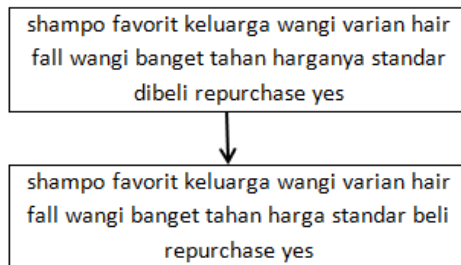


Gambar 3 Contoh Stopword Removal

### c. Stemming

Stemming adalah proses pembuangan imbuhan (affix), baik prefix maupun suffix, dari sebuah term untuk mendapatkan kata dasar (root atau stem) dari kata berimbuhan. Salah satu algoritma stemming untuk Bahasa Indonesia telah dikembangkan sebelumnya, yaitu algoritma Nazief & Adriani [19].

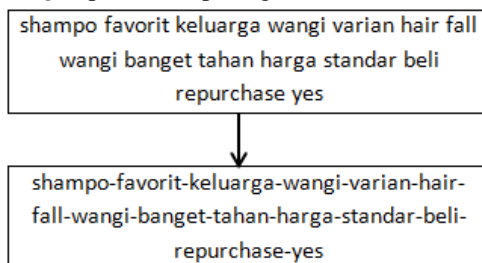
Algoritma Nazief & Adriani dikembangkan berdasarkan aturan morfologi Bahasa Indonesia yang mengelompokkan imbuhan menjadi awalan (prefix), sisipan (infix), akhiran (suffix) dan gabungan awalan akhiran (confixes).



Gambar 4 Contoh proses stemming

### d. Tokenizing

Tahap ini bertujuan untuk memecah kalimat menjadi per kata yang disebut dengan *token* [16]. Contoh proses *tokenizing* dapat dilihat pada gambar 4.



Gambar 5 Contoh tokenizing

## 6.3 Klasifikasi Naïve Bayes

Setelah dilakukan proses *preprocessing* selanjutnya data review diklasifikasikan menggunakan metode *Naïve Bayes Classifier* (NBC). Pada tahap ini dihitung probabilitas kemunculan kata pada data untuk menentukan skor rating sebuah review sebagai anggota kelas rating 1, 2, 3, 4, atau 5.

Proses klasifikasi dengan menggunakan metode NBC terdiri dari dua tahap yaitu tahap *training* dan *testing*. Tahapan klasifikasi dengan metode *Naïve Bayes Classifier* adalah sebagai berikut:

1. Menghitung nilai *prior* dari setiap kategori
2. Menghitung frekuensi setiap *term* pada *review* untuk setiap kategori
3. Menghitung nilai *likelihood* setiap *term* pada *review* untuk setiap kategori
4. Menghitung nilai *posterior* dari setiap *review* untuk setiap kategori

5. Menentukan klasifikasi *review* berdasarkan nilai *posterior* tertinggi.

## 6.4 Evaluasi

Pada tahap ini dilakukan pengujian dan evaluasi hasil pengukuran kualitas klasifikasi dengan membangun *confusion matrix*. Pengukuran evaluasi kinerja yang digunakan dalam penelitian ini adalah *accuracy*, *precision*, *recall*.

## 7. HASIL DAN PEMBAHASAN

Dalam penelitian ini, data yang digunakan untuk klasifikasi berjumlah 688 data review yang diambil dari *website femaledaily.com*. Data tersebut kemudian dibagi menjadi dua bagian dengan rasio 80:20 menjadi 550 untuk data *training* dan 138 untuk data *testing*. Persebaran data training sama rata berjumlah 110 pada setiap kelas. Persebaran jumlah dataset dapat dilihat pada tabel 2.

Tabel 2 Jumlah Dataset yang Digunakan

Dataset	R1	R2	R3	R4	R5	Total
Training	110	110	110	110	110	550
Test	19	26	28	30	35	138

Selanjutnya data kemudian diolah melalui tahap *text preprocessing* untuk membuat data lebih terstruktur. Setelah melalui tahap *preprocessing*, kumpulan *term* yang dihasilkan dari proses tersebut akan dilakukan proses klasifikasi dengan menggunakan metode *Naïve Bayes Classifier*. Untuk evaluasi hasil pengujian klasifikasi dapat dilihat pada tabel 3.

Tabel 3 Evaluasi Hasil Klasifikasi NBC

Percobaan ke-	Akurasi (%)	Recall (%)	Precision (%)
1	32,61	34,784	33,754
2	31,16	31,19	31,192
3	39,13	37,057	40,172
4	40,58	36,308	42,65
5	49,27	48,194	49,61
6	44,20	45,712	45,058
7	35,51	34,146	35,43
8	35,51	34,386	36,976
9	35,51	34,428	37,122
10	42,75	41,462	44,482
<b>Rata-rata</b>	<b>38,623</b>	<b>37,77</b>	<b>39,64</b>

Dari 10 kali percobaan, hasil terbaik diperoleh pada percobaan ke-5 dengan nilai akurasi mencapai 49,27% dengan nilai presisi 49,61% dan *recall* 48,19%.

**Tabel 4 Evaluasi hasil percobaan ke-5 dengan Confusion Matrix**

Predicted Class	R1	R2	R3	R4	R5	Class Recall (%)	
Actual Class	R1	15	1	0	3	0	78,95
	R2	10	8	1	3	4	30,77
	R3	4	12	3	7	3	10,71
	R4	3	3	1	16	6	53,33
	R5	2	4	1	2	26	74,29
Class Precision (%)	44,12	28,57	50,00	51,61	66,67	49,61 48,19	

Keterangan: R1 = Rating 1, R2 = Rating 2, R3 = Rating 3, R4 = Rating 4, R5 = Rating 5.

Hasil pengujian seperti pada tabel 4 menunjukkan jika total data yang diklasifikasi dengan benar sebanyak 68 data dari 138 data uji. Informasi lain yang didapatkan berdasarkan tabel *confusion matrix* meliputi empat parameter yaitu *class recall*, *class precision*, *error rate*, dan *accuracy*.

Hasil penghitungan *class recall* berdasarkan rumus (6) menghasilkan nilai tertinggi terdapat pada kelas rating 1 yang mencapai 78,95% dan nilai terendah pada rating 3 sebanyak 10,71%.

Sementara hasil penghitungan *class precision* berdasarkan rumus (5) didapatkan nilai tertinggi terjadi pada kelas rating 5 yang mencapai 66,67% dan nilai terendah terdapat pada kelas rating 2 dengan nilai 28,57%.

*Error rate* yang menunjukkan tingkat kejadian salah klasifikasi dihitung dengan rumus (7) sebagai berikut:

$$\text{Error Rate} = \frac{70}{138} \times 100\% = 50,72\%$$

*Accuracy* yang menunjukkan tingkat akurasi pada uji coba menggunakan metode NBC merupakan informasi utama dalam penelitian ini dapat dihitung dengan rumus (4) sebagai berikut:

$$\text{Accuracy} = \frac{15 + 8 + 3 + 16 + 26}{138} \times 100\% = 49,27\%$$

Klasifikasi prediksi rating review produk kosmetik dengan metode *Naïve Bayes Classifier* pada percobaan ke-5 penelitian ini mendapatkan tingkat akurasi sebesar 49,27%.

Berdasarkan tabel 4 *confusion matrix* menunjukkan rendahnya *class precision* kelas rating 2 jika dibandingkan kelas yang lainnya. Total jumlah *misclassification* pada kelas rating 2 mencapai 12 dari 28 data testing rating 2.

Kesalahan klasifikasi terbanyak terjadi pada kelas rating 2 yang salah terklasifikasi kedalam rating 3. Hal ini terjadi lantaran terdapat beberapa *term* yang memiliki probabilitas tinggi pada kelas 3, sehingga menyebabkan data terklasifikasi pada kelas tersebut.

Berdasarkan tabel 4 *confusion matrix*, nilai presisi jika di rata-rata adalah sebesar 48,19%, sementara rata-rata nilai *recall* adalah sebesar 49,61%. Kelas rating 1 dan rating 5 menunjukkan nilai *recall* di atas rata-rata dibandingkan dengan kelas lainnya. Hal ini terjadi karena data review kelas rating 1 dan rating 5 memiliki kata-kata yang spesifik sehingga lebih banyak data yang terklasifikasi dengan benar. Akan tetapi kelas rating 1 memiliki nilai presisi di bawah rata-rata. Terdapat data review R2, R3, R4, dan R5 yang terklasifikasi kedalam R1. Kesalahan klasifikasi tersebut terjadi karena pada data R2, R3, R4, dan R5 terdapat kata-kata yang memiliki kemiripan dengan kata-kata pada R1.

Dove Hair Fall Treatment Shampoo memiliki aroma yang khas dan membuat rileks. Kualitas untuk mencegah rambut rontok tidak terasa, bahkan kadang membuat rambut agak lepek. Berbagai produk dove hanya unggul di satu poin, yakni aromanya.

**Gambar 6 Contoh data Rating 2 yang salah terklasifikasi menjadi Rating 5**

Contoh pada gambar 6 Klasifikasi gagal karena dari fitur yang dihasilkan terdapat kata *khas*, *aroma*, *khas* dan *rambut*. Keempat kata tersebut memiliki jumlah frekuensi tertinggi pada kelas rating 5 sehingga mempengaruhi angka prediksi yang menjadi lebih tinggi di kelas tersebut. Hal ini menyebabkan sistem memprediksi teks review kedalam rating 5.

Jika dilihat berdasarkan nilai *recall*, secara keseluruhan sistem menunjukkan performa di atas rata-rata dalam mengklasifikasikan data review rating 1 dan 5. Meskipun pengklasifikasian pada data kelas rating 1 dan 5 menunjukkan performa demikian, akan tetapi masih terjadi kesalahan klasifikasi. Hal tersebut dapat dilihat seperti pada contoh kesalahan klasifikasi pada nomor 6, 7, dan 8.

Salah satu faktor yang menyebabkan performa pengklasifikasian data rating 1 dan 5 di atas rata-rata yaitu karena terdapat kata-kata yang menunjukkan perbedaan yang jelas antara sentimen positif atau negatif. Misalnya pada rating 1, terdapat kata-kata berkonotasi negatif yang memiliki frekuensi tertinggi pada kelas tersebut adalah kata *kering* dan *rontok*. Sementara pada rating 5 terdapat kata-kata berkonotasi positif seperti *wangi*, *love*, dan *repurchase*. Sementara pada data review rating 2, 3 dan 4 cenderung ditemukan kata-kata yang mirip sehingga terjadi banyak kesalahan klasifikasi pada data rating tersebut.



Penggunaan bahasa dalam teks review cenderung tidak baku atau informal dan rawan kesalahan penulisan. Di dalam beberapa data teks review terdapat beberapa kata yang mengandung sentimen yang mengalami kesalahan penulisan. Untuk itu dilakukan normalisasi pada kata-kata tersebut. Misalkan terjadi penulisan kata *enaaaak*, nantinya ditransformasi menjadi kata *enak* agar kata dapat diproses.

Inkonsistensi antara *user review* dengan *rating* juga terjadi. Misalkan teks review mengandung kata-kata negatif, akan tetapi rating yang diberikan oleh user adalah 4 atau sebaliknya. Hal ini juga menjadi salah satu penyebab terjadinya *misclassification* pada data review.

## 8. KESIMPULAN DAN SARAN

### 8.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dijalankan dapat diambil kesimpulan jika tingkat akurasi klasifikasi tertinggi mencapai 49,27% dengan akurasi rata-rata sebesar 38,62%. Berdasarkan nilai presisi secara keseluruhan sistem menunjukkan performa di atas rata-rata dalam mengklasifikasikan data review rating 1 dan 5.

### 8.2 Saran

Saran dalam penelitian ini adalah penelitian selanjutnya dapat melakukan pengambilan data yang lebih banyak jumlahnya untuk melihat konsistensi akurasi pada klasifikasi. Selain itu juga dapat ditambahkan fitur *Part-of-Speech Tagging* untuk mengetahui posisi sebuah kata dalam kalimat. Atau penerapan metode *rule based* juga dapat dicoba yaitu dengan membuat aturan kebahasaan untuk mengetahui susunan kata dalam kalimat, yang nantinya mampu menunjukkan tingkatan superlatif dan perubahan makna jika kata-kata tertentu disandingkan dengan kata negasi.

## 9. DAFTAR PUSTAKA

- [1] Liu, Bing. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers
- [2] Pamungkas, Dyarsa Singgih & Setiyanto, Noor Agung. (2015). Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci "KURIKULUM 2013".
- [3] Pang, Bo & Lee, Lillian. (2008). *Opinion Mining and Sentiment Analysis*. Foundation and Trends® in Information Retrieval. page 1-135
- [4] Asghar, Nabiha. (2016). *Yelp Dataset Challenge: Review Rating Prediction*
- [5] Chandani, V., Wahono, R.S., & Purwanto. (2015). *Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film*
- [6] Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). *Sentiment Classification of Internet Restaurant Reviews Written in Cantonese*. 38.
- [7] Fan, M., & Khademi, M. (2014). *Protecting a Business' Star in Yelp from Its Reviews Text Alone*.
- [8] Xu, Y., Wu, X., & Qinxia, W. (2014). *Sentiment Analysis of Yelp's Ratings Based on Text Reviews*.
- [9] Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approach in analyzing Unstructured Data*. New York: Cambridge University Press.
- [10] Ismaya, A. (2014). *Algoritma Ekstraksi Informasi Berbasis Aturan*. 03(04).
- [11] Larose, D. T. (2006). *Naïve Bayes Estimation and Bayesian Networks*, in *Data Mining Methods and Models*, John Wiley & Sons, Inc., Hoboken, NJ, USA, doi: 10.1002/0471756482.ch5
- [12] Ricci, F.; Rokach, L; et al. (2011). *Recommender Systems Handbook*. Berlin : Springer.
- [13] Arguello, J., (2013). *Naïve Bayes Text Classification* ([https://ils.unc.edu/courses/2013\\_fall/inls613\\_001/lectures/04NaiveBayesClassification.pdf](https://ils.unc.edu/courses/2013_fall/inls613_001/lectures/04NaiveBayesClassification.pdf). Diakses tanggal 31 Agustus 2018). The University of North California
- [14] Santra, A. K., & Christy, C. J. (2012). *Genetic Algorithm and Confusion Matrix for Document Clustering*.
- [15] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation," *Adv. Artif. Intell.*, vol. 4304, no. c, pp. 1015–1021, 2006.
- [16] Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*.
- [17] Dragut, E., Fang, F., Sistla, P., Yu, C., & Meng, W. (2009). *Stop Word and Related Problems in Web Interface Integration*.
- [18] Tala, F.Z., 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherlands: Institute for Logic, Language and Computation, Universiteit van Amsterdam.
- [19] Asian, J.(2007). *Effective Techniques for Indonesian Text Retrieval*.

