

Analisis Sentimen Pada *Review* Konsumen Menggunakan Metode *Naive Bayes* Dengan Seleksi Fitur *Chi Square* Untuk Rekomendasi Lokasi Makanan Tradisional

Novan Dimas Pratama¹, Yuita Arum Sari², Putra Pandu Adikara³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email : ¹novandimaspratama@gmail.com, ²yuita@ub.ac.id, ³adikara.putra@ub.ac.id

Abstrak

Ulasan konsumen pada sebuah restoran sangat berpengaruh dalam kualitas restoran itu sendiri. Banyak dari konsumen menuangkan kritik atau pendapatnya melalui media internet. Tujuan penelitian ini adalah untuk menganalisis sentimen opini dari konsumen makanan tradisional serta memberikan rekomendasi lokasi dengan kata kunci yang diinginkan. *Naive Bayes* adalah teknik *machine learning* yang sering digunakan untuk mengklasifikasikan data berupa teks. *Chi Square* merupakan seleksi fitur yang digunakan untuk menghitung tingkat dependensi sebuah fitur terhadap suatu kelas. Dalam penelitian ini, metode *Chi Square* memberikan nilai pada fitur yang selanjutnya diurutkan dan diseleksi menurut persentase yang diujikan. Fitur yang terseleksi digunakan untuk proses klasifikasi menggunakan metode *Naive Bayes*. Hasil akurasi klasifikasi dengan seleksi fitur 25% adalah sebesar 81%, dengan seleksi fitur 50% adalah sebesar 80 % dan dengan seleksi fitur 77% adalah sebesar 80%. Dari pengujian ini dapat disimpulkan bahwa seleksi fitur tidak begitu berpengaruh terhadap nilai hasil akurasi. Hal ini dapat diketahui perbedaan dari nilai akurasi antara menggunakan seleksi fitur dan tanpa menggunakan seleksi fitur yang tidak terlalu signifikan.

Kata kunci: ulasan konsumen, restoran makanan tradisional, analisis sentimen, *chi square*, *naive bayes*.

Abstract

Consumer reviews at a restaurant are very influential in the quality of the restaurant itself. Many of the consumers pour critics or opinions through the internet media. The goal of this research was to analyze the opinion sentiment from traditional food consumers as well as provide location recommendations with the desired keywords. *Naive Bayes* is a machine learning technique that is more used to classify text data. *Chi Square* is a feature selection used to calculate the level of a feature's dependencies on a class. In this study, *Chi Square* method gives value to the feature which is then sorted and selected according to percentage tested. Selected features are used for the classification process using the *Naive Bayes* method. The result of classification accuracy with 25% feature selection is 81%, with 50% feature selection is 80% and with 77% feature selection is 80%. From this test it can be concluded that feature selection is not so influential on the result value accuracy. It can be seen the difference of the accuracy value between using feature selection and without using a feature selection that is not very significant.

Keywords: consumer review, traditional food restaurant, sentiment analysis, *chi square*, *naive bayes*.

1. PENDAHULUAN

Bervariasinya makanan menjadi salah satu daya tarik bagi wisatawan dalam negeri maupun mancanegara. Mulai dari restoran mewah yang menjual makanan modernnya sampai kedai-kedai kecil pinggir jalan yang tak kalah dengan jajanan tradisionalnya menjadi acuan para pelancong untuk mengisi perut atau hanya sekedar ingin bersantai. Pada kota wisata yang

terkenal dengan buah apelnnya, terdapat banyak makanan-makanan tradisional yang diujakan oleh penjual untuk konsumen yang memang hobi dalam berwisata kuliner. Ketersediaan informasi yang ada memungkinkan konsumen untuk memberikan komentar, yang kemudian menjadi acuan bagi konsumen untuk melihat *review* dari restoran yang akan di kunjungi. Untuk restoran sendiri, dengan banyaknya informasi publik yang tersedia dapat digunakan

sebagai bahan pembelajaran untuk perkembangan restoran itu sendiri (Reyes & Rosso, 2012). Sebagian konsumen jajanan tradisional memang menuangkan ulasan kritik atau pendapat mereka melalui media sosial. Memang fitur komentar yang tersedia akan membantu seorang calon pengunjung dengan melihat *review* yang ada, akan tetapi dengan banyaknya serta bervariasinya komentar yang *netizen* berikan tentu akan memakan banyak waktu, dan apabila calon konsumen sedikit membaca ulasan yang ada, informasi yang ada akan menjadi bias.

Untuk menangani permasalahan tersebut dibutuhkan sebuah proses yang dapat memilah kata dari opini konsumen agar mudah menyeleksi tempat yang sekiranya cocok dan sesuai dengan selera. Analisis sentimen merupakan metode yang cocok untuk menyelesaikan permasalahan ini dengan cara mengumpulkan ulasan yang ada kemudian mengelompokkan ulasan tersebut menjadi suatu opini positif atau sebuah opini negatif. Analisis sentimen atau *opinion mining* digunakan untuk mendapatkan sebuah uraian umum tentang kualitas sebuah layanan, apakah layanan tersebut cenderung mendapatkan nilai positif, negative maupun netral. (Istiqomah, Maharani & Adiwijaya, 2014).

Penelitian sebelumnya telah dilakukan oleh Romi Dalam menganalisis sentimen dari *review* restoran menggunakan metode *AdaBoost* dan *Naive Bayes* serta *Information Gain* untuk seleksi fiturnya. Metode *AdaBoost* digunakan untuk meningkatkan tingkat akurasi lemah yang dilakukan algoritma sebelumnya (Wahono & Utami, 2015).

Untuk metode pengklasifikasian sendiri banyak peneliti menggunakan *Naive Bayes* dimana sebuah teks akan diklasifikasikan dalam *machine learning* berdasarkan probabilitas (Zhang & Gao, 2011). Klasifikasi digunakan untuk penggolongan kata pada data ulasan yang telah didapatkan, termasuk di kelas manakah sebuah opini tersebut. *Naive Bayes* merupakan teknologi *pre-processing* dalam klasifikasi fitur, yang menambah skalabilitas, akurasi dan efisiensi yang tentunya sangat dalam proses klasifikasi sebuah teks. Sebagai alat klasifikasi, *Naive Bayes* dianggap efisien dan sederhana, serta sensitif pada seleksi fitur (Chen, et al., 2009).

Tahapan lain yang kerap ditemui pada klasifikasi teks ialah seleksi fitur. Seleksi fitur dapat menjadikan alat klasifikasi lebih baik dan

lebih efisien, serta efektif dengan cara menyeleksi jumlah fitur data latih, serta menentukan fitur yang cocok untuk dipertimbangkan dalam proses pembelajaran (Wahono & Utami, 2015). Terdapat dua metode utama pada seleksi fitur dalam *machine learning*. Akurasi klasifikasi dari beberapa algoritma sebagai evaluator digunakan pada metode *Wrapper*. Metode filter tersusun dari dokumen *frequency*, *ChiSquare*, *mutual information*, dan *information gain* (Chen, et al., 2009).

Beberapa metode filter yang sering digunakan adalah *Information Gain* dan *Chi Square*. Namun metode *Chi Square* sering digunakan untuk pengklasifikasian dengan dokumen yang banyak dibandingkan *Information Gain*. Semakin banyak jumlah dokumen yang digunakan dalam klasifikasi dapat meningkatkan nilai *F-Measure* klasifikasi teks dengan metode *Chi Square*, dan menurunkan nilai *F-Measure* klasifikasi teks dengan metode *Information Gain* (Sofiana, Astatina & Suryani, 2012).

Chi Square juga digunakan dalam klasifikasi teks sebagai seleksi fitur dalam penelitian dilakukan oleh Juen, Kencana dan Tjokorda. *Chi Square* berfungsi untuk menyeleksi fitur dengan menguji independensi suatu kata dengan kelasnya. Contoh penggunaan seleksi fitur adalah untuk menghilangkan fitur yang tidak digunakan dalam proses klasifikasi (Ling, Kencana & Oka, 2014).

Berdasarkan permasalahan di atas maka penulis memutuskan untuk memberikan rekomendasi lokasi makanan tradisional berdasarkan *review* konsumen, maka digunakanlah metode *Naive Bayes* dengan seleksi fitur *Chi Square*. Dengan metode tersebut, diharapkan akan menghasilkan klasifikasi dokumen dengan nilai akurasi yang tinggi.

2. METODOLOGI PENELITIAN

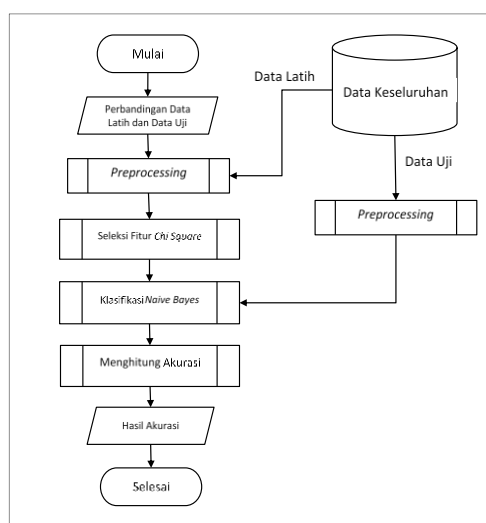
2.1 Pengumpulan Data

Data di dapatkan pada website www.tripadvisor.co.id. Penulis menggunakan 200 ulasan yang berasal dari konsumen restoran makanan tradisional pada Kota Malang. Lokasi yang digunakan tidak menyeluruh, hanya sebagian tempat yang digunakan dan terdapat ulasan yang bersifat positif atau negatif. Ulasan-

ulasan yang digunakan-terdiri dari 100 komentar positif dan 100 komentar negatif. Kategori positif atau negatif didapatkan dari nilai yang diberikan pengulas, jika nilai yang diberikan 4 dan atau 5 maka termasuk dalam kategori positif. Sebaliknya, jika nilai yang diberikan 2 dan atau 1 maka termasuk dalam kategori negatif. Pemberian data *keyword* berdasarkan menu makanan apa yang disajikan oleh restoran tersebut.

2.2 Perancangan Sistem

Pada sub-bab ini di jelaskan mengenai diagram alir algoritme metode *Naïve Bayes* dengan seleksi fitur *Chi Square* yang digunakan pada sistem. Diagram alir merupakan salah satu metode untuk mendeskripsikan suatu algoritme atau proses pada simbol-simbol yang memudahkan pengguna untuk memahami proses dalam suatu sistem. Berikut adalah Diagram alir analisis sentimen *review* konsumen menggunakan metode *Naive Bayes* dengan seleksi fitur *Chi Square* ditunjukkan pada Gambar 1.



Gambar 1. Perancangan Sistem

2.2.1 Preprocessing

Pada pemrosesan teks dilakukan tiga tahap dalam memproses teks menjadi lebih terstruktur. Salah satu tahapan pada pemrosesan teks ialah *preprocessing*. Tahap ini dimana data mentah diolah menjadi data yang digunakan untuk proses klasifikasi (Hadna, Santosa & Winarno, 2016).

Pada pemrosesan teks, *preprocessing* dibagi dalam 3 langkah, yaitu *Tokenization*, *Stopword Removal* dan *Stemming*. *Tokenization* adalah pemotongan string input tiap kata

penyusunnya. *Stopword Removal* merupakan proses penyeleksian kata-kata penting dari hasil fitur, yang nantinya digunakan untuk proses selanjutnya. *Stopword list* yang digunakan adalah *stopword list* berbahasa Indonesia yang berasal dari Tala. *Stemming* ialah proses pencarian *stem-base* dari suatu kata. *Stemmer* yang digunakan adalah *PHP library* dari *stemmer Sastrawi* berbahasa Indonesia yang berasal dari *website github.com/sastrawi*. *Stemmer Sastrawi* ini menggunakan *composer* untuk menjalankannya.

2.2.2 Metode Chi Square

Chi Square merupakan metode yang digunakan untuk menghitung tingkat dependensi fitur. Dalam pemrosesan teks, biasanya dua kelas digunakan untuk mengukur tingkat ketergantungan antara sebuah kata t dan kelas tertentu c antara kedua tabel tersebut (Suharno, Fauzi & Perdana, 2017). Fungsi dari *Chi Square* dapat dilihat dari Persamaan (1).

$$x^2(t, c) = \frac{N(AD - CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

Keterangan:

t = Term

c = Kelas/Kategori

N = Jumlah data latih

A = Jumlah dokumen pada kelas c yang memuat t

B = Jumlah dokumen bukan kelas c yang memuat t

C = Jumlah dokumen pada kelas c yang tidak memuat t

D = Jumlah dokumen bukan kelas c yang tidak memuat t

Untuk mendapatkan nilai *Chi Square* dari suatu kata pada sebuah kelas, dibutuhkan nilai *Chi Square* tunggal dari kata tersebut. Agar mendapatkan nilai *Chi Square* tunggal dari suatu kata dapat menjumlahkan nilai *Chi Square* dari masing-masing kelas. Nilai *Chi Square* tunggal tiap kata di dapatkan melalui algoritme pada Persamaan (2).

$$X^2(t) = \sum_{c=1}^k x^2(t, c) \quad (2)$$

Setelah diketahui, dilakukan pengurutan kata berdasarkan nilai *Chi Square* tertinggi hingga terendah. Hal ini menandakan bahwa semakin besar nilai *Chi Square*, semakin

dependen suatu fitur, dan semakin penting fitur tersebut untuk digunakan dalam proses klasifikasi.

2.2.3 Metode Naive Bayes

Naive Bayes adalah sebuah proses klasifikasi yang menggunakan perhitungan probabilitas. Metode *Naive Bayes* berdasar pada algoritme dengan teknik klasifikasi yang telah melalui uji efisiensi dan efektifitas dalam suatu *database* dengan jumlah data yang besar, *Naive Bayes* mempunyai tingkat akurasi dan efisiensi yang tinggi (Kusumadewi, 2009). Dasar dari teorema Bayes dinyatakan dalam Persamaan (3).

$$c_{NB} = \operatorname{argmax} P(c) \prod_{x \in X} P(x|c) \quad (3)$$

Keterangan:

- c_{NB} = Kelas hasil penghitungan *Naive Bayes*.
 $P(c)$ = Probabilitas kelas pada data latih.
 $P(x|c)$ = Peluang fitur x muncul pada dokumen c .
 $x \in X$ = jumlah fitur unik pada dokumen.

Probabilitas parameter $P(c_j)$ dijelaskan pada Persamaan (4).

$$P(c_j) = \frac{N_j}{N} \quad (4)$$

Keterangan:

- $P(c_j)$ = Probabilitas masing-masing kelas pada data latih.
 N_j = Jumlah dokumen pada suatu kelas pada data latih.
 N = Jumlah dokumen pada data latih.

Kelemahan dari probabilitas parameter $P(x|c)$ apabila bernilai nol. Oleh karena itu digunakan metode *smoothing* pada *Naive Bayes* untuk menyelesaikan permasalahan tersebut. Metode yang digunakan pada penelitian ini adalah metode *Laplacian Smoothing* (Hafilizara, 2014). Metode tersebut dijelaskan pada Persamaan (5).

$$P(x_i|c_j) = \frac{\text{count}(x_i, c_j) + 1}{(\sum_{w \in V} \text{count}(x, c_j)) + |V|} \quad (5)$$

Keterangan:

$|V|$ = Banyaknya fitur pada data keseluruhan.

2.2.4 Akurasi

Akurasi adalah penghitungan dari perbandingan antara jumlah data dokumen yang relevan dan jumlah keseluruhan dokumen dalam *database* (Hasugian, 2016). Persamaan akurasi dapat dilihat pada Persamaan (6).

$$\text{Akurasi} = \frac{\text{Jumlah Dokumen Relevan}}{\text{Jumlah keseluruhan Dokumen}} \quad (6)$$

3. PENGUJIAN DAN ANALISIS

3.1 Pengujian akurasi

Tujuan dilakukannya pengujian akurasi adalah untuk melihat kinerja sistem untuk menganalisis sentimen pada *review* konsumen menggunakan metode *Naive Bayes* dengan seleksi fitur *Chi Square*. Pengujian dibedakan menjadi empat tahap, tahap tersebut disesuaikan pada persentase hasil seleksi fitur data latih dengan nilai *Chi Square*. Tahap pertama, dengan seleksi fitur data latih dengan nilai *Chi Square* sebesar 25%. Tahap kedua, dengan seleksi fitur data latih dengan nilai *Chi Square* sebesar 50%. Tahap ketiga, dengan seleksi fitur data latih dengan nilai *Chi Square* sebesar 75%. Tahap keempat, dengan seleksi fitur data latih dengan nilai *Chi Square* sebesar 100%.

3.1.1 Seleksi Fitur Persentase 25%

Hasil pengujian tahap pertama dengan persentase seleksi fitur 25% ditunjukkan pada Tabel 1.

| | Tabel 1. Fitur Data Latih 25% | |
|------------------|-------------------------------|----------------------|
| | Ground Truth positif | Ground Truth negatif |
| Prediksi positif | 36 | 5 |
| Prediksi negatif | 14 | 45 |

Hasil pengujian tahap pertama didapatkan data dengan prediksi benar kelas positif sebanyak 36 dokumen dan prediksi salah kelas positif sebanyak 14 dokumen. Prediksi benar kelas negatif sebanyak 45 dokumen dan prediksi salah kelas negatif sebanyak 5 dokumen. Penghitungan nilai akurasi didapat dengan menggunakan Persamaan (6). Nilai akurasi yang didapat yaitu:

$$\text{Nilai Akurasi} = \frac{36+45}{36+14+5+45} = 81\%$$

3.1.2 Seleksi Fitur Persentase 50 %

Hasil pengujian tahap pertama dengan persentase seleksi fitur 50% ditunjukkan pada Tabel 2.

Tabel 2. Fitur Data Latih 50%

| | Ground Truth positif | Ground Truth negatif |
|------------------|----------------------|----------------------|
| Prediksi positif | 35 | 5 |
| Prediksi negatif | 15 | 45 |

Hasil pengujian tahap pertama didapatkan data dengan prediksi benar kelas positif sebanyak 35 dokumen dan prediksi salah kelas positif sebanyak 15 dokumen. Prediksi benar kelas negatif sebanyak 45 dokumen dan prediksi salah kelas negatif sebanyak 5 dokumen. Penghitungan nilai akurasi dapat menggunakan pada Persamaan 6. Nilai akurasi yang diperoleh adalah:

$$\text{Nilai Akurasi} = \frac{35+45}{35+15+5+45} = 80\%$$

3.1.3 Seleksi Fitur Persentase 75 %

Hasil pengujian tahap pertama dengan persentase seleksi fitur 75% ditunjukkan pada Tabel 3.

Tabel 3 Fitur Data Latih 75%

| | Ground Truth positif | Ground Truth negatif |
|------------------|----------------------|----------------------|
| Prediksi positif | 32 | 5 |
| Prediksi negatif | 18 | 45 |

Hasil pengujian tahap pertama didapatkan data dengan prediksi benar kelas positif sebanyak 32 dokumen dan prediksi salah kelas positif sebanyak 18 dokumen. Prediksi benar kelas negatif sebanyak 45 dokumen dan prediksi salah kelas negatif sebanyak 5 dokumen. Penghitungan nilai akurasi dapat menggunakan Persamaan 6. Nilai akurasi yang diperoleh adalah:

$$\text{Nilai Akurasi} = \frac{32+45}{31+19+6+44} = 77\%$$

3.1.4 Seleksi Fitur Persentase 100 %

Hasil pengujian tahap pertama dengan persentase seleksi fitur 100% ditunjukkan pada Tabel 4.

Tabel 4 Fitur Data Latih 100%

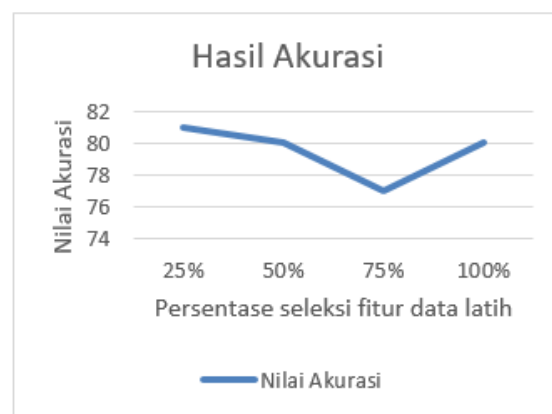
| | Ground Truth Positif | Ground Truth Negatif |
|------------------|----------------------|----------------------|
| Prediksi positif | 34 | 4 |
| Prediksi negatif | 16 | 46 |

Hasil pengujian tahap pertama didapatkan data dengan prediksi benar kelas positif sebanyak 35 dokumen dan prediksi salah kelas positif sebanyak 15 dokumen. Prediksi benar kelas negatif sebanyak 45 dokumen dan prediksi salah kelas negatif sebanyak 5 dokumen. Penghitungan nilai akurasi dapat menggunakan Persamaan 2.7. Nilai akurasi yang diperoleh adalah:

$$\text{Nilai Akurasi} = \frac{34+46}{34+16+4+46} = 80\%$$

3.2 Analisis pengujian akurasi

Dari pengujian akurasi yang dilakukan, dapat dilihat hasil analisis pada Gambar 2.



Gambar 2. Grafik hasil akurasi

Jumlah nilai akurasi selalu berubah-ubah pada tiap persentase seleksi fitur data latih. Perbedaan hasil akurasi disebabkan oleh seleksi fitur yang dilakukan dengan menggunakan metode *Chi Square*. Nilai *Chi Square* dari fitur diurutkan dari tinggi ke rendah untuk mendapatkan fitur yang akan digunakan sesuai dengan perkalian persentase. Terdapat hasil nilai yang sama pada persentase seleksi fitur 50% dan 100%. Persentase seleksi fitur 25% merupakan nilai akurasi tertinggi dengan nilai 81%. Sedangkan persentase fitur 75% merupakan nilai akurasi terendah dengan nilai 77%. Persentase fitur 100% yang juga merupakan klasifikasi tanpa seleksi fitur mendapatkan hasil akurasi dengan nilai 80%.

Dari pengujian ini dapat disimpulkan bahwa seleksi fitur tidak begitu berpengaruh terhadap nilai hasil akurasi. Hal ini dapat diketahui dari perbedaan nilai akurasi antara menggunakan seleksi fitur dan tidak, yang menghasilkan nilai yang tidak terlalu signifikan.

3.3 Hasil rekomendasi

Tujuan dilakukannya hasil rekomendasi adalah untuk melihat rekomendasi lokasi makanan tradisional yang ada pada data dan ditampilkan oleh sistem. Pengujian ini didapat dari hasil klasifikasi kelas data uji pada sistem. Nantinya sistem akan menampilkan beberapa lokasi berdasarkan persamaan kata kunci dan kata dalam dokumen yang telah dimasukkan sebelumnya. Sistem juga akan menampilkan *value* dari perbandingan sentimen positif dan negatif. *Value* didapat dari jumlah dokumen dengan sentimen positif pada dikurangi jumlah dokumen dengan sentimen negatif pada lokasi tertentu. Berikut adalah pengujian rekomendasi dengan kata kunci “bakso” pada Tabel 5.

Tabel 5. Rekomendasi Lokasi

| No. | Lokasi | Value |
|-----|---------------------------|-------|
| 1 | Bakso Bakar Pak Man | 8 |
| 2 | Bakwan Subur | 1 |
| 3 | Bakso Bakar Pahlawan Trip | -1 |
| 4 | Bakso Bakar Trowulan | -1 |
| 5 | Bakso Gondhol | -1 |
| 6 | Bakso Kota Cak Man | -1 |
| 7 | Bakso President | -9 |

4. PENUTUP

4.1 Kesimpulan

Berdasarkan perolehan hasil dari perancangan, implementasi dan pengujian yang dilakukan maka bisa disimpulkan bahwa dalam menganalisis sentimen menggunakan metode *Naive Bayes* berdasarkan *review* konsumen dengan menggunakan seleksi fitur *Chi Square* untuk penentuan tempat makanan tradisional yang pertama perlu dilakukan adalah menentukan fitur dari data latih yang akan digunakan dalam pengujian lewat *Preprocessing* teks. Fitur yang dihasilkan selanjutnya akan diurutkan dan diseleksi sesuai dengan tingkat dependensinya menggunakan metode *Chi Square*, penyeleksian dilakukan dengan perkalian hasil persentase dengan fitur data latih. Selanjutnya dilakukan klasifikasi dokumen menggunakan metode *Naive Bayes* dengan menghitung nilai *prior* dan *posterior* untuk mendapatkan kelas dari data uji.

Nilai akurasi dari data uji didapatkan setelah menghitung jumlah dokumen dengan kelas benar dibagi dengan jumlah dokumen keseluruhan. Jumlah data uji yang digunakan dalam pengujian adalah 100 *review* konsumen

yang terdiri dari 50 *review* positif dan 50 *review* negatif. Pada pengujian pertama menggunakan persentase 25% dari seleksi fitur didapatkan nilai akurasi sebesar 81%. Pada pengujian kedua menggunakan persentase 50% seleksi fitur didapatkan nilai akurasi sebesar 80%. Pada pengujian ketiga menggunakan persentase 75% dari seleksi fitur didapatkan nilai akurasi sebesar 77%. Pada pengujian terakhir tanpa menggunakan seleksi fitur didapatkan nilai akurasi sebesar 80%.

4.2 Saran

Berikut ini adalah saran dari penulis yang dapat digunakan dalam penelitian selanjutnya. Peneliti dapat menggunakan pakar untuk menentukan data apakah termasuk dalam kelas positif atau negatif. Perbaikan kata dalam dokumen sangat dibutuhkan untuk mendapatkan tingkat dependensi fitur dalam suatu kelas. Sebab, jika terdapat kata *slang* dalam suatu dokumen akan mempengaruhi nilai keterkaitan dari fitur atau kata tersebut. Penggunaan *bigrams* atau *N-grams* dapat membantu dalam penyeleksian fitur dan menghasilkan nilai akurasi yang tinggi.

5. DAFTAR PUSTAKA

- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). *Feature Selection for Text Classification with Naïve Bayes*. *Expert Systems with Applications*, 36(3), 5432–5435.
- Feldman, R. 2013. *Techniques and Applications for Sentiment Analysis*. *Communications of the ACM*, 56(4), 82.
- Hadna, N. M. S., Santosa, P. I. & Winarno, W. W., 2016. Studi Literatur tentang Perbandingan Metode untuk Proses Analisis Sentimen di Twitter.
- Hafilizara, M., 2014. Metode Smoothing dalam Naive Bayes untuk Klasifikasi Email Spam.
- Hasugian, J., 2016. *Penelusuran Informasi Ilmiah Secara Online: Perlakuan terhadap Seorang Pencari Informasi sebagai Real User*. s.l.:Jurnal Studi Perpustakaan dan Informasi.
- Istiqomah, S. M., Maharani, W. & Adiwijaya, 2014. *Opinion Mining pada Twitter Menggunakan Klasifikasi Sentimen pada Hashtag berbasis Graf*.

- Kang, H., Yoo, S. J. & Han, D., 2012. *Senti-lexicon and Improved Naive Bayes Algorithms for Sentiment Analysis of Restaurant Reviews*. s.l.:s.n.
- Kusumadewi, S., 2009. Klasifikasi Status Gizi menggunakan Naive Bayesian Classification.
- Ling, J., Kencana, I. P. E. N. & Oka, T. B., 2014. *Analisis Sentimen Menggunakan Metode Naive Bayes Classifier dengan Seleksi Fitur Chi Square*.
- Liu, B., 2012. *Sentiment Analysis and Subjectivity*. Chicago: Department of Computer Science.
- Muthia, D. A., 2017. *Analisis Sentimen pada Review Restoran dengan Teks Bahasa Indonesia menggunakan Algoritma Naive Bayes*.
- Sofiana, I., Astatina, I. & Suryani, A. A., 2012. *Analisis Pengaruh Feature Selection menggunakan Information Gain dan Chi Square untuk Kategorisasi Teks Berbahasa Indonesia*.
- Suharno, C. F., Fauzi, M. A. & Perdana, R. S., 2017. Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online menggunakan Metode K-Nearest Neighbors (K-NN) dan Chi Square.
- Wahono, R. S. & Utami, L. D., 2015. *Integrasi Metode Information Gain untuk Seleksi Fitur dan AdaBoost untuk Mengurangi Bias pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes*.
- Zhang, W., & Gao, F. 2011. *An Improvement to Naive Bayes for Text Classification*. *Advanced in Control Engineering and Information Science*, 15, 2160–2164.