

Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura

Vynska Amalia Permadi

Program Studi Informatika, Fakultas Teknik Industri,
Universitas Pembangunan Nasional “Veteran” Yogyakarta
Jl. Babarsari 2, Tambak Bayan, Depok, Kab. Sleman, 55281, Propinsi DI.
Yogyakarta, Indonesia
Email: vynspermadi@upnyk.ac.id

Abstract. *Sentiment Analysis Using Naive Bayes Classifier Against Restaurant Reviews in Singapore. Various restaurant options bring up a problem for diners to pick a restaurant to dine in. Thus, visitors usually perceive the restaurant's recommendation or rating in advance to know other diners' opinions about the restaurant. Previous restaurant diners' comments can be presented in sentiment analysis to determine their satisfaction. This research investigates the Naïve Bayes Classifier algorithm's performance in classifying visitors' sentiment based on restaurant diner comments. We will group visitors' comments into two types of sentiment: positive (satisfied) and negative (unsatisfied). The results of the data classification test are analyzed to determine its accuracy. The grouping of visitor satisfaction reviews using the naïve bayes algorithm provides an accuracy score of 73%. Besides, we visualize the research classification results in the browser-based R Shiny web application through word cloud and diagrams.*

Keywords: *restaurant review, sentiment analysis, Naïve Bayes Classifier*

Abstrak. *Variasi pilihan restoran yang tidak sedikit menjadi salah satu masalah bagi pengunjung ketika ingin memilih restoran. Sehingga, pengunjung biasanya melihat rekomendasi atau penilaian pengunjung lain terhadap restoran tersebut terlebih dahulu untuk mengetahui penilaian pengunjung lain terhadap restoran tersebut. Penilaian atau review pengunjung dapat disajikan dalam analisis sentimen berdasarkan komentar para pengunjung restoran sebelumnya untuk melihat kepuasan pengunjung terhadap restoran tersebut. Penelitian ini dilakukan untuk mengetahui performa algoritma Naïve Bayes Classifier dalam melakukan klasifikasi sentimen berdasarkan komentar pengunjung restoran. Penelitian dilakukan dengan mengklasifikasikan data komentar pengunjung restoran menjadi dua kategori sentimen, yaitu: positif (satisfied) dan negatif (unsatisfied). Hasil pengujian pengklasifikasian data kemudian dianalisis akurasi. Hasil pengelompokan review kepuasan pengunjung menggunakan algoritma naïve bayes memberikan nilai akurasi sebesar 73%. Visualisasi hasil klasifikasi dari analisis kemudian ditampilkan pada aplikasi berbasis web yaitu R Shiny berupa wordcloud dan diagram.*

Kata Kunci: *penilaian restoran, analisis sentimen, Naïve Bayes Classifier*

1. Pendahuluan

Wisata kuliner merupakan hal yang disukai wisatawan saat mengunjungi suatu daerah atau tempat wisata baik di dalam maupun di luar negeri. Dengan banyaknya variasi kuliner yang ada, maka semakin banyak pula penjual makanan atau restoran yang menyediakan berbagai macam jenis kuliner. Jumlah variasi restoran yang banyak juga menjadi satu masalah bagi pengunjung ketika akan memilih restoran mana yang akan mereka kunjungi, sehingga umumnya calon pengunjung akan menentukan pilihannya berdasarkan rekomendasi atau penilaian pengunjung lain terhadap restoran tersebut. Dengan melakukan analisis sentimen terhadap penilaian atau *review* dari pengunjung sebelumnya, dapat dilakukan identifikasi kepuasan pengunjung terhadap restoran tersebut.

Analisis sentimen telah banyak dilakukan dengan berbagai metode, objek, data maupun parameter. Pada penelitian yang dilakukan oleh Sipayung dkk. [1], telah dirancang suatu sistem analisis sentimen dari komentar pelanggan terhadap produk dan layanan hotel dengan menggunakan metode *Naïve Bayes Classifier* yang menghasilkan 155 komentar kategori positif dan 20 komentar kategori negatif dari 175 data latih dan diperoleh pula nilai akurasi sebesar 75,42%. Selanjutnya, pengujian analisis sentimen juga dilakukan dengan metode lain, yaitu penggabungan algoritma *Naïve Bayes* dan *Particle Swarm Optimization* (PSO) [2] untuk menganalisa sentimen *review* restoran. Dengan menggabungkan dua algoritma tersebut, dihasilkan tingkat akurasi yang cukup baik, yaitu sebesar 88,50%. Pada penelitian oleh Muthia [3], selain menggunakan algoritma *Naïve Bayes*, diterapkan pula metode pemilihan fitur menggunakan algoritma Genetika yang terbukti dapat meningkatkan akurasi klasifikasi sehingga tingkat akurasi yang diberikan adalah sebesar 90,50%. Pengukuran akurasi pada penelitian tersebut diberikan dalam bentuk *confusion matrix* dan kurva ROC.

Berbagai metode analisis sentimen memiliki kelebihan dan kekurangan yang beragam. Penelitian ini bertujuan untuk menganalisis komentar pelanggan restoran dengan metode yang *Naïve Bayes Classifier*. Metode ini dipilih karena seperti yang disebutkan oleh Imron [4] bahwa penggunaan metode *Naïve Bayes Classifier* hanya membutuhkan jumlah *data training* yang kecil sebagai parameter dalam proses klasifikasi.

2. Tinjauan Pustaka

Saat ini, telah banyak situs, produk, maupun perusahaan yang mengidentifikasi maupun menganalisis pendapat atau opini penggunanya dengan melakukan analisis sentimen. Pada penelitiannya Liu [5] mengatakan bahwa analisis sentimen dilakukan dalam bentuk studi komputasi linguistik untuk menganalisis pendapat, ekspresi, maupun perasaan penilai terhadap berbagai macam objek tulisan atau teks, seperti: entitas, individu, topik, maupun bentuk obyek lainnya. Terkait dengan studi sentimen yang dilakukan, hasil analisis pendapat tersebut selanjutnya juga dapat dijadikan oleh individu maupun perusahaan untuk menentukan kebijakan atau keputusan tindak lanjut [6]. Secara spesifik, dalam implementasinya di dunia bisnis, analisis sentimen atau *option mining* yang merupakan cabang dari *text mining* seringkali digunakan untuk menganalisis opini pelanggan terhadap produk dan layanan suatu bisnis [7].

Penelitian yang berkaitan dengan analisis sentimen sering dilakukan karena topik ini merupakan salah satu topik yang menarik untuk dibahas dalam penelitian. Sebagai contoh, pada penelitian [8], Ratnawati membahas mengenai pengklasifikasian opini film berdasarkan komentar terhadap film tersebut yang disampaikan pada media sosial *Twitter*. Analisis sentimen pada penelitian tersebut dilakukan dengan menggunakan metode *deep learning* dengan akurasi yang didapat sebesar 90%. Falahah dan Nur [9] juga telah melakukan pengklasifikasian opini pada media sosial *Twitter* untuk menganalisis sentimen masyarakat dengan menggunakan algoritma *Naïve Bayes* untuk menganalisis opini mengenai layanan pemerintah. Dengan metode yang sama, yaitu *Naïve Bayes classifier*, Sunni dan Widyantoro [10] juga telah melakukan analisis sentimen terhadap opini mengenai tokoh publik yang disampaikan pada media sosial *Twitter*. Selain *Naïve Bayes*, penelitian tersebut juga menambahkan penggunaan metode TF-IDF untuk pengekstrakan topik.

3. Metodologi Penelitian

Proses penelitian dilakukan dengan menggunakan salah satu metode *machine learning*, yaitu *supervised learning* yang menurut Sinaga [11], merupakan salah satu metode pembelajaran mesin yang membutuhkan data pelatihan atau *data testing* dan juga data latihan atau *data training* sehingga kemudian hasilnya adalah berupa *output* sesuai dengan proses *training data* yang telah dilakukan. Metode *supervised learning* yang dipilih untuk analisis sentimen adalah *Naïve Bayes classifier*. Analisis sentimen dilakukan untuk menentukan kategori komentar positif dan negatif.

3.1. Pengumpulan Data

Pengambilan data uji dilakukan dengan metode *web scraping* pada salah satu situs penyedia *review* pengguna terhadap destinasi wisata, hotel, restoran, dan lainnya, yaitu Tripadvisor (<https://www.tripadvisor.com/>), khususnya *review* mengenai restoran yang ada di Singapura. Data yang akan digunakan untuk proses analisis sentimen adalah nama restoran dan tautan *review* pengunjung terhadap restoran. Untuk mengukur akurasi pengelompokan sentimen, digunakan *data training* dari data klasifikasi “*Restaurant Reviews*” yang tersedia di situs Kaggle (<https://www.kaggle.com/hj5992/restaurantreviews>). Data dari situs tersebut terdiri dari dua atribut, yaitu komentar dan hasil analisis sentimen (direpresentasikan dalam biner 0 untuk sentimen negatif dan 1 untuk sentimen positif).

3.2. Preprocessing

Langkah ini dilakukan untuk mempermudah proses analisis data. Pada langkah ini dilakukan beberapa tahapan. Beberapa tahapan yang dilakukan dalam *preprocessing* diantaranya: menghilangkan tanda baca atau *punctuation*, serta *hashtag* ataupun *mention*, mengubah semua teks menjadi huruf kecil (*lower case*), menghapus *stop word*, mengubah kata yang terdapat imbuhan menjadi kata dasar (*stemming*), serta mengubah kata yang tidak baku menjadi kata baku.

3.3. Klasifikasi Sentimen dengan Algoritma Naïve Bayes

Algoritma *Naïve Bayes* merupakan salah satu bentuk pengklasifikasian dengan mempertimbangkan nilai probabilitas statistik. Sehingga, prediksi keanggotaan kelas suatu data baru dilakukan dengan memprediksi probabilitas keanggotaan dari data tersebut [12]. Persamaan teorema *Bayes* dinyatakan pada Persamaan 1.

$$P(c|X) = \frac{P(X | c).P(c)}{P(X)} \tag{1}$$

Keterangan:

$P(c|x)$: Probabilitas hipotesis berdasar kondisi (*posteriori probability*)

$P(c)$: Probabilitas hipotesis (*prior probability*)

$P(x|c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

Penelitian ini akan mengimplementasikan aturan *Bayes* pada studi kasus analisis sentimen restoran. Aturan *Bayes* yang digunakan dinyatakan pada Persamaan 2 berikut:

$$P(cj|X) = \frac{p(X | cj).p(cj)}{p(X)} \tag{2}$$

Dimana cj adalah kategori teks yang akan diklasifikasikan dan $p(cj)$ merupakan probabilitas prior dari kategori teks cj . Sedangkan d merupakan dokumen teks yang direpresentasikan sebagai himpunan kata ($W1, W2, \dots Wn$), dimana $W1$ adalah kata pertama, $W2$ adalah kata kedua dan seterusnya. Pada saat proses pengklasifikasian dokumen teks, maka pendekatan *Bayes* akan memilih kategori yang memiliki probabilitas paling tinggi (C_{MAP}) pada Persamaan 3, yaitu:

$$C_{MAP} = \operatorname{argmax} \frac{p(cj).p(X | cj)}{p(X)} \tag{3}$$

Nilai $p(X)$ dapat diabaikan karena nilainya adalah konstan untuk semua c_j , sehingga Persamaan 4 dapat dituliskan:

$$C_{MAX} = \operatorname{argmax} p(c_j) \cdot p(X|c_j) \quad (4)$$

Probabilitas $p(c_j)$ dapat diestimasi dengan menghitung jumlah dokumen training pada setiap kategori c_j . Sedangkan untuk menghitung distribusi $p(X | c_j)$ akan sulit karena jumlah *term* menjadi sangat besar. Hal ini disebabkan jumlah term tersebut sama dengan jumlah semua kombinasi posisi kata dikalikan dengan jumlah kategori yang akan diklasifikasikan. Dengan pendekatan *Naïve Bayes* yang mengasumsikan bahwa setiap kata dalam setiap kategori adalah tidak bergantung satu sama lain, maka perhitungan dapat lebih disederhanakan dan dapat dituliskan seperti Persamaan 5 sebagai berikut:

$$P(X|c_j) = \prod_{i=1}^n p(w_j | c_j) \quad (5)$$

Dengan menggunakan Persamaan 2, maka Persamaan 5 dapat dituliskan menjadi Persamaan 6 sebagai berikut:

$$C_{MAP} = \operatorname{argmax} p(c_j) \cdot \prod_{i=1}^n p(w_j | c_j) \quad (6)$$

Nilai $p(c_j)$ dan $p(w_i | c_j)$ dihitung pada saat proses pelatihan dimana persamaannya diberikan pada Persamaan 7 dan Persamaan 8 sebagai berikut :

$$P(c_j) = \frac{|docs_j|}{|contoh|} \quad (7)$$

$$P(w_j|c_j) = \frac{1 + n_i}{|C| + n(kosakata)} \quad (8)$$

- $p(w_i | c_j)$: Probabilitas kata w_i pada kategori
- $c_j | docs_j |$: Jumlah dokumen pada kategori
- $j | contoh |$: Jumlah seluruh dokumen sampel yang digunakan dalam proses *training*
- n_i : Frekuensi kemunculan kata w_i pada kategori
- $c_j | C |$: Jumlah semua kata pada kategori
- $c_j n(kosakata)$: Jumlah kata yang unik pada semua *data training*.

3.4. Analisis Sentimen dengan *Confusion Matrix*

Pada tahap ini dilakukan analisis hasil dari pengklasifikasian sentimen berdasarkan hasil *actual* dan *prediction* untuk *review* restoran. Dengan menampilkan hasil perbandingan antara prediksi dan aktual, maka dapat mengetahui perbedaan hasil pengklasifikasian sentimen antara prediksi dan aktual pada data *review* restoran.

Penghitungan *confusion matrix* menunjukkan jumlah prediksi benar dan juga prediksi salah terhadap kelas negatif maupun kelas positif. Pada penelitian ini, karena jenis klasifikasi yang digunakan adalah klasifikasi berupa positif maupun negatif, maka *confusion matrix*-nya dapat disajikan seperti pada Tabel 1 [12].

Tabel 1. *Confusion Matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan:

- TP adalah *True Positive*, yaitu jumlah data yang nilai aktualnya adalah positif dan terklasifikasi sebagai positif.
- TN adalah *True Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif dan terklasifikasi sebagai negatif.
- FN adalah *False Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif namun terklasifikasi salah.
- FP adalah *False Positive*, yaitu jumlah data yang nilai aktualnya adalah positif namun terklasifikasi salah.

Setelah mengetahui jumlah nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP), dapat dilakukan perhitungan selanjutnya untuk mencari nilai akurasi, presisi dan *recall*. Nilai akurasi memberikan informasi berupa seberapa akurat data terklasifikasi dengan benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar terhadap keseluruhan jumlah data. Nilai akurasi dapat diperoleh dengan melakukan perhitungan pada Persamaan 9. Nilai presisi menggambarkan jumlah data dengan hasil aktual positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasikan sebagai positif [13]. Untuk memperoleh nilai presisi, dapat dilakukan perhitungan dengan Persamaan 10. Sementara itu, nilai *recall* menggambarkan keberhasilan pengklasifikasian, yaitu dengan menghitung persentase data kategori positif yang terklasifikasikan dengan benar [13]. Nilai *recall* diperoleh melalui Persamaan 11.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

3.5. Visualisasi Analisis Sentimen dengan R Shiny

Visualisasi dari hasil analisis sentimen ditampilkan pada *Shiny* yang merupakan salah satu *framework* dalam R untuk membangun aplikasi berbasis *website*. Untuk menampilkan *shiny* maka harus digunakan Bahasa pemrograman R. Menampilkan visualisasi data dengan *Shiny* sangat mudah, dan dapat menghasilkan halaman *website* yang lebih interaktif. Visualisasi ditampilkan dalam dua bentuk, yaitu *word cloud* dan histogram. *Word cloud* merupakan tampilan dari kata-kata yang sering muncul, dan besar kecilnya menunjukkan frekuensi kemunculannya dalam data. Sehingga kata terbanyak akan ditampilkan dalam ukuran yang lebih besar dan sebaliknya, tampilan kata akan semakin kecil jika jumlah frekuensi kemunculannya sedikit. Histogram merupakan hasil dari frekuensi data yang ditampilkan dalam bentuk grafik batang yang dapat disajikan dalam berbagai variasi warna. Setiap batang menunjukkan masing-masing kata yang sering muncul terhadap jumlah frekuensi kemunculannya. Pada histogram, tampilan diagram batang yang semakin tinggi diberikan untuk data yang frekuensinya semakin banyak, sebaliknya, ukuran batang akan semakin pendek jika frekuensinya semakin sedikit.

4. Hasil dan Diskusi

Pada penelitian ini, data *review* restoran diklasifikasikan sentimennya dengan algoritma *Naïve Bayes classifier* dan berikut adalah hasil dan pembahasan dari langkah uji coba yang dilakukan pada penelitian:

4.1. Pengambilan dan Pengumpulan Data

Pengambilan *data testing* atau data uji dilakukan dengan metode *web scraping* menggunakan *package* “*rvest*” pada bahasa pemrograman R. Untuk menggunakan *package* tersebut, tentukan terlebih dahulu *url website* yang akan diambil informasinya. Berdasarkan hasil *web scraping* yang dilakukan pada satu halaman Tripadvisor, diperoleh 33 nama restoran dan halaman tautan *review* restoran pada setiap nama restoran. Hasil pada langkah ini dicontohkan pada Tabel 2.

Tabel 2. Contoh Data Hasil Web Scraping pada Halaman Situs TripAdvisor

No	Nama Restoran	Halaman Tautan Review
1	Botany	/Restaurant_Review-g294265-d16855342-Reviews-Botany-Singapore.html
2	1. Le Noir Bar & Lounge	/Restaurant_Review-g294265-d18190065-Reviews-Le_Noir_Bar_Lounge-Singapore.html
3	2. Entre-Nous creperie	/Restaurant_Review-g294265-d1193730-Reviews-Entre_Nous_creperie-Singapore.html
4	3. Alma By Juan Amador	/Restaurant_Review-g294265-d8504143-Reviews-Alma_By_Juan_Amador-Singapore.html
5	4. Fu Lin Men (CSC)	/Restaurant_Review-g294265-d13873832-Reviews-Fu_Lin_Men_CSC-Singapore.html
6	5. Fu Lin Men Chinese Restaurant (SRC)	/Restaurant_Review-g294265-d15608624-Reviews-Fu_Lin_Men_Chinese_Restaurant_SRC-Singapore.html

Data training yang digunakan sebagai data latih untuk menentukan klasifikasi *review* restoran terdiri dari 1000 data yang berisi contoh kata dan juga kelas sentimen berupa angka 0 yang menunjukkan kategori negatif dan angka 1 yang menunjukkan kategori positif. Data ini diperoleh dari *dataset* “*Restaurant Review*” yang diperoleh dari situs Kaggle. Pada Tabel 3, diberikan contoh *data training* yang digunakan.

Tabel 3. Contoh Data Training "Restaurant Review"

No	Review	Sentimen
1	Wow... Loved this place	1
2	Crust is not good	0
3	Not tasty and the texture was just nasty	0
4	Stopped by during the late May bank holiday off Rick Steve recommendation and loved it	1
5	The selection on the menu was great and so were the prices	1

4.2. Preprocessing

Cleaning data nama restoran dilakukan dengan penghapusan karakter angka dengan pemrograman R, sebelum data melalui proses *cleaning* bentuk data perlu diubah terlebih dahulu dari *data frame* menjadi *data corpus*. Setelah dilakukan proses *cleaning*, data kemudian disimpan dalam *data frame* baru. Proses *cleaning* yang dilakukan memanfaatkan beberapa *package* yang tersedia dalam R, seperti “*corpus*” yang berguna untuk merubah data menjadi *data corpus*. Dengan telah dilakukannya proses *cleaning*, nama restoran telah terformat rapi (tidak ada lagi tanda baca maupun karakter yang kurang tepat) seperti yang ditunjukkan pada pada Tabel 4 di bawah ini.

Tabel 4. Contoh Hasil Preprocessing Terhadap Data Hasil Web Scraping

No	Nama Restoran	Halaman Tautan Review
1	Botany	/Restaurant_Review-g294265-d16855342-Reviews-Botany-Singapore.html
2	Le Noir Bar Lounge	/Restaurant_Review-g294265-d18190065-Reviews-Le_Noir_Bar_Lounge-Singapore.html
3	EntreNous creperie	/Restaurant_Review-g294265-d1193730-Reviews-Entre_Nous_creperie-Singapore.html
4	Alma by Juan Amador	/Restaurant_Review-g294265-d8504143-Reviews-Alma_By_Juan_Amador-Singapore.html

No	Nama Restoran	Halaman Tautan Review
5	Fu Lin Men CSC	/Restaurant_Review-g294265-d13873832-Reviews-Fu_Lin_Men_CSC-Singapore.html
6	Fu Lin Men Chinese Restaurant SRC	/Restaurant_Review-g294265-d15608624-Reviews-Fu_Lin_Men_Chinese_Restaurant_SRC-Singapore.html

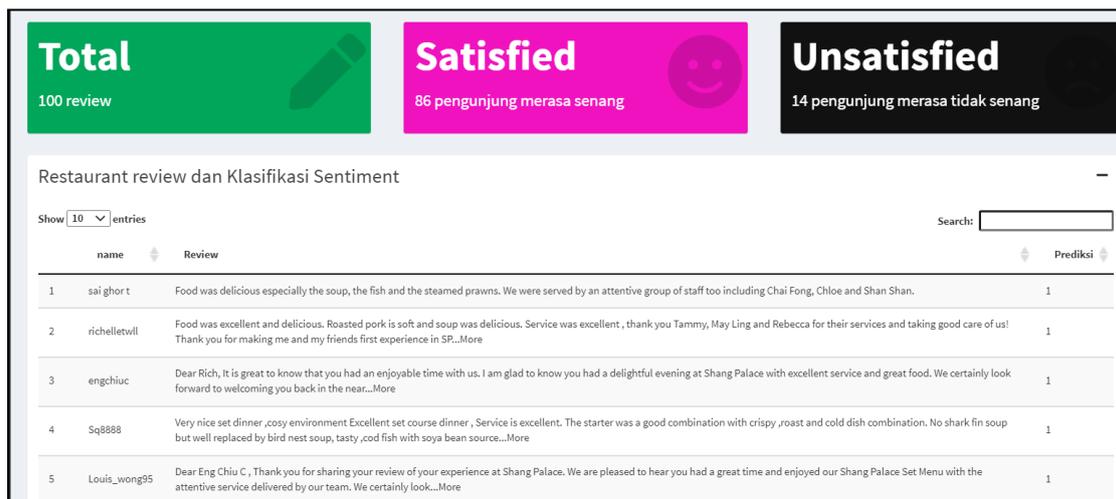
Setelah pengambilan data nama restoran dan link *review* dari restoran maka data selanjutnya yang akan diambil yaitu data *review* dan juga *reviewer* dari setiap restoran dengan mengambil link *review* yang ada pada data sebelumnya. Data yang didapatkan berupa *review* atau komentar terkait dengan restoran dan juga *reviewer* atau nama dari *user* yang memberikan komentar, seperti yang ditunjukkan pada Tabel 5. Dari setiap restoran memiliki puluhan hingga ratusan *review* seperti pada restoran yang pertama yaitu “Botany” memiliki data *review* sebanyak 33 *list* beserta dengan data *review* dan berikut tabel data dari *review* dan *reviewer*:

Tabel 5. Contoh Data Review dari Para Pengunjung

No	Isi Review	Reviewer
1	a nice little cafe at robertson quay. ordered takeaway of pasta and juice. will be back soon to try other dishes on the menu	bbbbbw
2	Love the lunch that we had here. Really taste great. And moreover, is pet friendly and love the place alot.	SweetAlisya
3	Came here for brunch friends as we had a puppy with us, and Botany allows is a pet-friendly cafe. Food was decent, and the interior of the restaurant is pretty. Came back a second time with anther group of friends in just 1 month.	Irene S
4	The interior and exterior of this cafe is really pretty. Great atmosphere. I had the mushroom soup, but it isn't even like a soup. It contains so much mushroom that I cannot even drink anything. It is very dry just looks like a side of...More	pandaat501
5	Excellent Tap room, quiet place to hang out with friends. As its little away frm Clark Quayso its kind of less crowded. However, weekends are packed in a way. Is located on riverbank, across Warehouse Hotel and serene surroundings.	Kilroy_Mariner

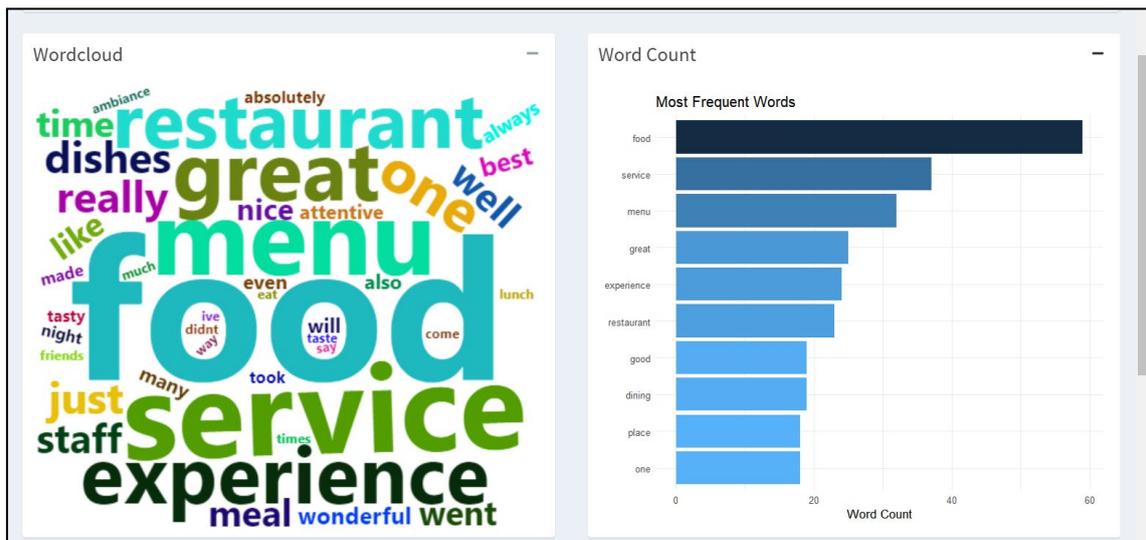
4.3. Visualisasi Hasil Analisis pada Shiny

Hasil visualisasi data ditampilkan dalam *shinyApp*, seperti yang ditunjukkan pada Gambar 1. Pada gambar tersebut ditunjukkan bahwa telah terdapat 100 data *review* yang telah diklasifikasikan dari salah satu restoran Singapura. Berdasarkan gambar tersebut, dapat dilihat bahwa hasil dari klasifikasi 0 dikelompokkan sebagai *unsatisfied* dan 1 sebagai *satisfied*.



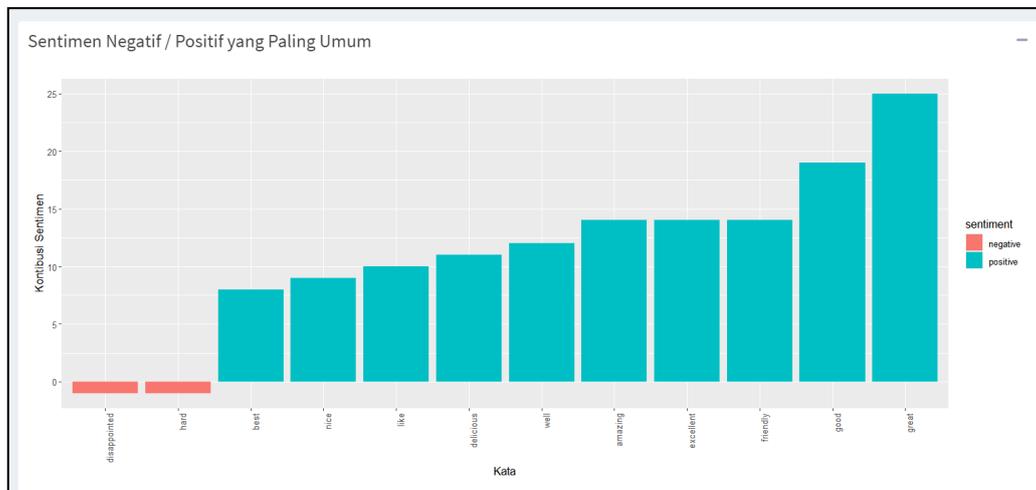
Gambar 1. Dashboard Klasifikasi Review Restoran

Gambar 2 menunjukkan tampilan *wordcloud* dari *review* restoran dan *word count* (Frekuensi 10 kata yang paling sering muncul). Berdasarkan tampilan *wordcloud*, dapat dilihat bahwa kata “*food*” adalah kata yang sering muncul pada komentar restoran.



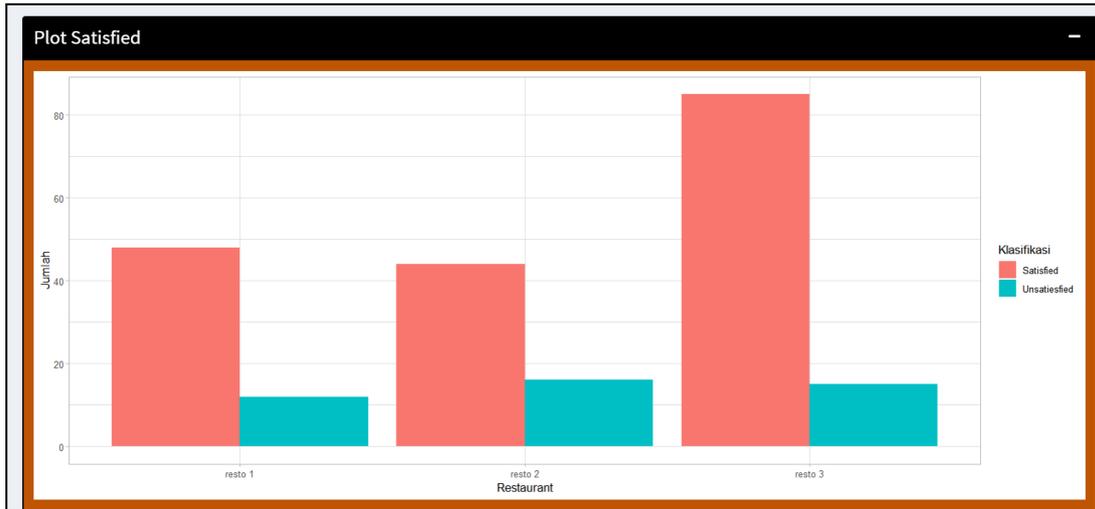
Gambar 2. Tampilan Hasil Analisis berdasarkan *Wordcloud* dan *Word Count*

Gambar 3 menunjukkan tampilan histogram dari *website shiny* yang juga menunjukkan analisis frekuensi dari pengklasifikasian *review* restoran Singapura. Klasifikasi ditunjukkan dalam dua warna yang berbeda, sentimen positif dengan warna biru dan sentimen negatif diberikan dalam warna merah. Kata-kata yang ditampilkan dalam histogram adalah kata-kata sentimen positif dan sentimen negatif yang paling sering muncul, diantaranya: *disappointed* dan *hard* untuk sentiment negatif, sedangkan kata-kata positif yang sering muncul adalah: *best*, *nice*, *like*, *delicious*, *well*, *amazing*, *excellent*, *friendly*, *good*, dan *great*.



Gambar 3. Histogram *Sentimen Positif dan Negatif*

Tampilan perbandingan analisis *review* sentimen terhadap restoran ditunjukkan pada Gambar 4. Pada bagian ini, ditampilkan frekuensi sentimen *satisfied* dan *unsatisfied* terhadap restoran yang direview.



Gambar 4. Diagram Perbandingan Frekuensi *Satisfied* dan *Unsatisfied* pada *Review Sentimen Restoran*

4.4. Analisis Naïve Bayes Classifier

Hasil pengklasifikasian sentimen ditunjukkan pada Gambar 5. Apabila nilai *class* (nilai aktual) sama dengan hasil prediksi maka *review* tersebut telah berhasil diklasifikasikan dengan benar, sebaliknya pengklasifikasian dianggap belum berhasil diidentifikasi dengan benar.

text	class	prediksi
197 On the good side, the staff was genuinely pleasant and enth...	1	1
76 Delicious NYC bagels, good selections of cream cheese, real ...	1	1
163 You get incredibly fresh fish, prepared with care.	1	0
540 The pan cakes everyone are raving about taste like a sugary ...	0	0
835 My sashimi was poor quality being soggy and tasteless.	0	0
628 Things that went wrong: - They burned the saganaki.	0	1
340 The food is good.	1	1
895 The service was terrible, food was mediocre.	0	0
442 Perfect for someone (me) who only likes beer ice cold, or in ...	1	0
736 I won't try going back there even if it's empty.	0	0
732 Terrible management.	0	0
201 This is an Outstanding little restaurant with some of the Best...	1	0
34 The cocktails are all handmade and delicious.	1	1
168 In the summer, you can dine in a charming outdoor patio - s...	1	1
343 An absolute must visit!	1	0
883 My fella got the huevos rancheros and they didn't look too ...	0	0
144 The restaurant atmosphere was exquisite.	1	1
592 Must have been an off night at this place.	0	1

Gambar 5. Hasil Prediksi dan Aktual Review Restoran

Berdasarkan penelitian, diperoleh *review* restoran pada *class* 0 dan diklasifikasikan dengan benar sebanyak 111 *review*. Sedangkan hasil klasifikasi salah pada *class* 0 adalah sebanyak 41 *review*. Untuk hasil klasifikasi benar pada *class* 1 terdiri dari 109 *review*, sedangkan hasil klasifikasi salah pada *class* 1 adalah 39 *review*. Berdasarkan hasil tersebut, dapat disusun *confusion matrix* nya seperti yang ditunjukkan pada Tabel 6. Jumlah prediksi yang terklasifikasikan secara benar pada *class* 0 adalah 111 *review*, yang selanjutnya diidentifikasi sebagai TP (*true positive*) dan jumlah prediksi yang terklasifikasi salah pada *class* 0 adalah 41 selanjutnya disebut sebagai FP (*false positive*). Kemudian, jumlah prediksi yang terklasifikasi dengan benar pada *class* 1 adalah 109 disebut TN (*true negative*) dan jumlah

prediksi yang terklasifikasi salah pada *class* 1 adalah 39 yang kemudian disebut sebagai FN (*false negative*).

Tabel 6. Hasil *Confusion Matrix* Dari *Review Restoran*

Predictions	Actual Reviewer	
	0	1
0	111	41
1	39	109

Berdasarkan *confusion matrix* tersebut, selanjutnya dilakukan perhitungan nilai *recall* dengan menggunakan rumus pada Persamaan 11.

$$Recall = \frac{TP}{TP + FN} = \frac{111}{111 + 39} = 0.74$$

Setelah itu dilakukan pula perhitungan nilai *precision* dengan menggunakan rumus pada Persamaan 10:

$$Precision = \frac{TP}{TP + FP} = \frac{111}{111 + 41} = 0.7302631578947368$$

Terakhir, nilai akurasi diperoleh dengan melakukan perhitungan menggunakan rumus pada Persamaan 9:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{111 + 109}{111 + 109 + 41 + 39} = 0.7333333333333333$$

Dari hasil perhitungan tersebut, dapat disimpulkan bahwa berdasarkan hasil perhitungan *recall*, sebanyak 74% data terklasifikasikan dengan benar, dimana *data review* yang diprediksi bernilai negatif adalah benar bernilai negatif dan sebaliknya. Selanjutnya, nilai *precision* digunakan untuk mengidentifikasi rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif, dan diperoleh hasil sebesar 73,02%. Tingkat akurasi yang diberikan pada model dalam penelitian ini mencapai 73,33% yang berarti bahwa 73,33% *data review* diprediksi dengan benar pada semua *class* (*class* 0 atau sentimen negatif dan *class* 1 atau sentimen positif).

5. Kesimpulan dan Saran

Penelitian untuk menganalisis sentimen terhadap data restoran yang ada di Singapura dengan algoritma *Naïve Bayes classifier* yang telah dilakukan menghasilkan kesimpulan bahwa algoritma *Naïve Bayes classifier* dapat menentukan klasifikasi dari suatu *review* ke dalam dua kategori yaitu positif yang diimplementasikan dalam kata *satisfied* atau puas dan juga *unsatisfied* atau tidak puas. Berdasarkan hasil pengujian yang dilakukan, pengklasifikasian sentimen dengan algoritma *naïve bayes* memberikan nilai *precision* sebesar 73,02%, *recall* sebesar 74%, serta akurasi sebesar 73,33%. Untuk hasil yang lebih maksimal, maka analisis dapat dilakukan dengan data yang lebih baik atau lebih bersih dengan cara menambahkan proses yang lebih mendetail lagi pada tahapan *preprocessing* yang dilakukan. Dapat pula digunakan jumlah data yang lebih banyak maupun penggabungan dengan metode lain agar dapat diperoleh hasil akurasi yang lebih baik.

Referensi

- [1] E. M. Sipayung, H. Maharani dan I. Zefanya, "Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode Naive Bayes Classifier," *Jurnal Sistem Informasi (JSI)*, vol. 8, no. 1, pp. 958-965, Apr. 2016.
- [2] K. Solecha, "Analisa sentimen dengan algoritma Naïve Bayes Classifier berbasis particle swarm optimization untuk review restoran," *Journal Speed-Sentra Penelitian Engineering dan Edukasi*, vol. 11, pp. 34-41, Feb. 2019.
- [3] D. Muthia, "Analisis sentimen pada review restoran dengan teks bahasa Indonesia menggunakan algoritma Naive Bayes," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 2, no. 2, pp. 39-45, Feb. 2017.
- [4] A. Imron, "Analisis sentimen terhadap tempat wisata di kabupaten Rembang menggunakan metode Naive Bayes Classifier," *Dspace Universitas Islam Indonesia*, Feb. 27, 2019. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/14268>
- [5] A. Sinaga, "Implementasi sentiment analysis untuk menentukan tingkat popularitas tujuan wisata," in *Seminar Nasional Teknologi dan Rekayasa Informatika*, Malang, Indonesia, Nov. 24 – 25, 2017, pp. 12-18.
- [6] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, New York, USA: Springer, 2011.
- [7] A. Akbar, E. Sedyono dan O. Nurhayati, "Analisis sentimen berbasis Ontologi di level kalimat untuk mengukur persepsi produk," *Jurnal Sistem Informasi Bisnis*, vol. 2, pp. 84-97, Jul. 2015.
- [8] I. F. Rozi, S. H. Pramono dan E. A. Dahlan, "Implementasi opinion mining (analisis sentimen) untuk ekstraksi data opini publik pada perguruan tinggi," *Jurnal EECCIS*, vol. 6, no. 1, pp. 37-43, Jun. 2012.
- [9] F. Ratnawati, "Implementasi algoritma Naive Bayes terhadap analisis sentimen opini film pada Twitter," *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 3, no. 1, pp. 50-59, Jun. 2018.
- [10] Falahah dan D. A. Nur, "Pengembangan aplikasi sentiment analysis menggunakan metode Naïve Bayes (studi kasus sentiment analysis dari media Twitter)," in *Seminar Nasional Sistem Informasi Indonesia*, Surabaya, Indonesia, Nov. 2-3, 2015, pp. 335-340.
- [11] I. Sunni dan D. H. Widyantoro, "Analisis sentimen dan ekstraksi topik penentu sentimen pada opini terhadap tokoh publik," *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, vol. 1, no. 2, pp. 200-206, Jul. 2012.
- [12] F. Handayani dan S. Pribadi, "Implementasi algoritma Naive Bayes Classifier dalam pengklasifikasian teks otomatis pengaduan dan pelaporan masyarakat melalui layanan call center 110," *Jurnal Teknik Elektro*, vol. 7, no. 1, pp. 19-24, Jan. 2015.
- [13] M. Sokolova dan G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, Jul. 2009.