

# Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran Singapura

Cici Yuriza<sup>1</sup>, Alivi Milova<sup>2</sup>

Jurusan Teknik Informatika, Fakultas Teknik Industri  
Universitas Pembangunan Nasional “Veteran” Yogyakarta  
Jl. Babarsari 2 Yogyakarta 55281 (Kampus Unit II)

<sup>1</sup>ciciyuriza999@gmail.com , <sup>2</sup>alivimilova7@gmail.com

## Abstrak

Restoran merupakan salah satu tempat yang dikunjungi untuk menikmati beragam kuliner yang ada di suatu daerah atau tempat tertentu. Jumlah restoran yang tidak sedikit menjadi salah satu masalah bagi pengunjung ketika memilih restoran sehingga diperlukan rekomendasi atau penilaian pengunjung lain terhadap restoran tersebut, dan penilaian atau review pengunjung dapat dilakukan dengan analisis sentiment berdasarkan komentar para pengunjung restoran untuk melihat kepuasan pengunjung terhadap restoran tersebut. Analisis sentiment dilakukan dengan berbagai metode seperti Support Vector Machine (SVM), Maximum Entropy, Naïve Bayes, dan metode lainnya. Penelitian ini dilakukan untuk mengetahui performa algoritma Naïve Bayes Classifier dalam melakukan klasifikasi berdasarkan komentar pengunjung restoran. Dengan data komentar pengunjung dan data review restoran sebagai parameter data latih maka menghasilkan dua kategori yaitu positif diimplementasikan dengan kata *satisfied* dan negative yang diimplementasikan dengan kata *unsatisfied* kemudian menghasilkan nilai akurasi sebesar 75 % dalam analisis sentiment menggunakan metode Naïve Bayes Classifier.

**Kata Kunci :** Restoran, Review, Analisis Sentimen, Naïve Bayes Classifier.

## I. PENDAHULUAN

Wisata kuliner merupakan hal yang disukai ketika mengunjungi suatu daerah atau tempat wisata baik di dalam maupun di luar negeri, dengan banyaknya kuliner yang ada maka tidak sedikit pula tempat penjualan makanan atau restoran yang menyediakan berbagai macam makanan saat berlibur. Dan restoran yang banyak menjadi satu masalah bagi pengunjung Ketika diharuskan memilih restoran yang akan mereka kunjungi sehingga diperlukan rekomendasi atau penilaian pengunjung lain terhadap restoran tersebut, dan penilaian atau review pengunjung dapat dilakukan dengan analisis sentiment berdasarkan komentar para pengunjung restoran untuk melihat kepuasan pengunjung terhadap restoran tersebut.

Analisis sentimen telah banyak dilakukan dengan berbagai metode dan dengan banyak obyek juga berbagai data parameter, seperti yang dilakukan oleh (Sipayung et al., 2016) dimana merancang system analisis sentimen dari komentar pelanggan terhadap produk dan layanan

hotel dengan menggunakan metode Naïve Bayes Classifier menghasilkan 155 komentar kategori positif dan 20 komentar kategori negatif dari 175 data latih dengan nilai akurasi 75.42 %. Selanjutnya pengujian analisis sentiment juga dilakukan dengan metode lain seperti yang dilakukan oleh (Solecha, 2019) yaitu analisa sentimen terhadap review restoran dilakukan dengan algoritma Naïve Bayes berbasis Particle Swarm Optimization yang menghasilkan tingkat akurasi sebesar 88,50 %. Kemudian berbagai metode alternatif juga dilakukan untuk meningkatkan performa dari analisis sentimen seperti dilakukan oleh (Muthia, 2017) pengujian dilakukan dengan menerapkan metode pemilihan fitur menggunakan algoritma genetika yang bisa meningkatkan akurasi dari klasifikasi Naïve Bayes dan hasil yang didapat menunjukkan tingkat akurasi yang besar yaitu 90.50% yang diukur dengan *confusion matrix* dan kurva ROC.

Berbagai metode memiliki kelebihan dan kekurangan yang berbeda dan pada penelitian ini dilakukan analisis sentiment terhadap komentar pelanggan restoran

dengan metode yang dipilih yaitu Naïve Bayes Classifier karena seperti yang disebutkan oleh (Imron, 2019) bahwa penggunaan metode Naïve Bayes Classifier hanya membutuhkan jumlah data training yang kecil sebagai parameter dalam klasifikasi.

## II. METODE PENELITIAN

Proses penelitian proyek menggunakan Teknik data mining yaitu *Supervised Learning* yang dijelaskan oleh (Sinaga, 2017) bahwa *Supervised Learning* merupakan metode belajar dimana terdapat pelatih atau data testing dan juga latihan atau data training yang kemudian menjadi output dari proses training tersebut. Kemudian dari hasil data training dan data testing dihasilkan analisis sentiment dengan menerapkan Algoritma Naïve Bayes Classifier. Analisis sentiment dilakukan untuk menentukan kategori komentar positif dan negative.

### 2.1 Pengumpulan data

Data testing diperoleh dari website <https://www.tripadvisor.com> khususnya dari restoran yang ada di Singapura. Dengan pengambilan dua field yaitu nama *restoran* dan link *review* restoran, sedangkan untuk data training menggunakan parameter yang sudah tersedia di website <https://www.kaggle.com> untuk review restoran untuk menentukan kategori positif dan negative. Berikut contoh data yang diambil dan digunakan

### 2.2 Preprocessing

Preprocessing dilakukan mulai dari proses cleaning untuk mempermudah proses analisis seperti menghilangkan tanda baca atau *punctuation* dan juga *hashtag* atau *mention*. Kemudian proses merubah semua huruf menjadi huruf kecil (*tolower*) dan terakhir yaitu proses menghapus *stopword* atau kata yang tidak sesuai dengan kaidah Bahasa Indonesia, mengubah kata yang terdapat imbuhan menjadi kata dasar (*stemming*), mengubah kata yang mempunyai makna menjadi hanya sebuah kata (*lemmatisasi*) dan terakhir mengubah kata yang *typo* menjadi kata baku (*slankword*) sehingga setelah preprocessing

selesai maka data sudah dapat diolah.

### 2.3 Proses Klasifikasi

Proses klasifikasi dengan metode supervised learning dilakukan dengan beberapa algoritma seperti KNN, Support Vector Machine dan juga Naïve Bayes. Dan analisis sentimen dari model supervised learning akan menghasilkan klasifikasi positif, negatif atau netral. Algoritma Naïve Bayes merupakan teknik yang diterapkan untuk menentukan kelas dari setiap masalah atau obyek yang sudah dibagi berdasarkan setiap masalah, naïve bayes dipilih menjadi salah satu algoritma yang banyak digunakan dalam proses analisis sentimen karena sederhana, cepat dan juga memiliki tingkat akurasi yang tinggi seperti yang dijelaskan oleh (Hanggara et al., 2017).

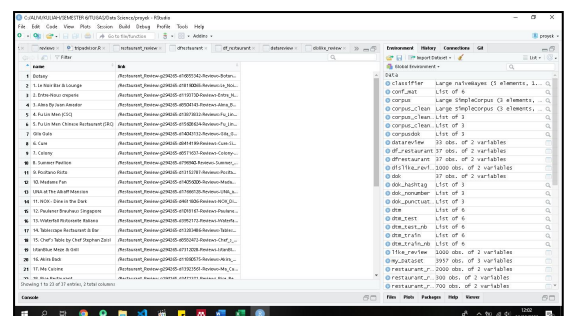
## III. HASIL DAN PEMBAHASAN

Pada penelitian proyek ini data diolah dengan algoritma Naïve Bayes Classifier dan berikut adalah hasil dan pembahasan dari Langkah uji coba yang dilakukan pada penelitian :

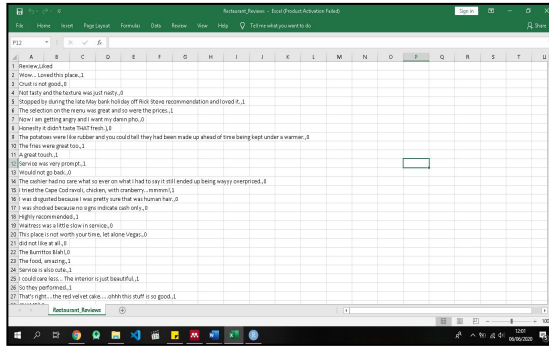
### 3.1 Pengambilan dan pengolahan data

#### 1. Pengambilan data dengan scraping web

Dari *page* pertama data yang dihasilkan memiliki 33 nama restoran dengan link review pada setiap nama restoran dan kemudian terdapat 1000 data pada data training untuk review restoran, berikut adalah tampilan dari data scraping dan juga data training



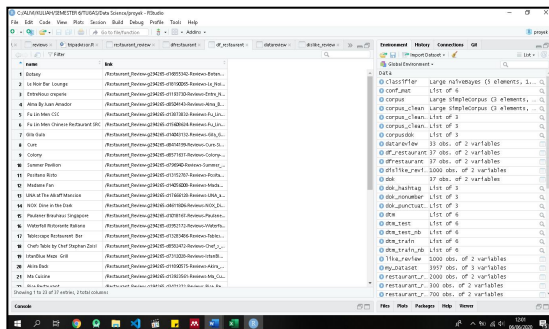
Gambar 1 data scraping



Gambar 2 data training

## 2. Cleaning data nama restoran

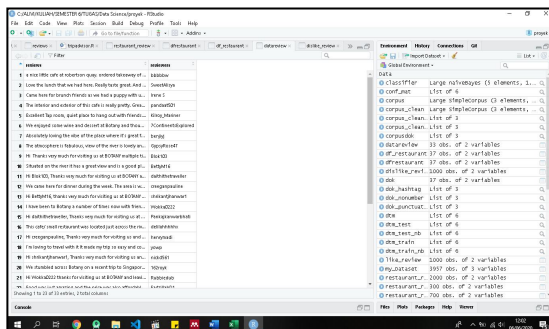
Hasil pengambilan data nama restoran memiliki karakter angka yang dapat menghambat proses implementasi analisis sehingga dilakukan cleansing pada data nama dengan menghapus karakter nomor, dan berikut tampilan dari data nama yang sudah di bersihkan :



Gambar 3 cleaning data scraping

## 3. Data Review dan reviewer restoran

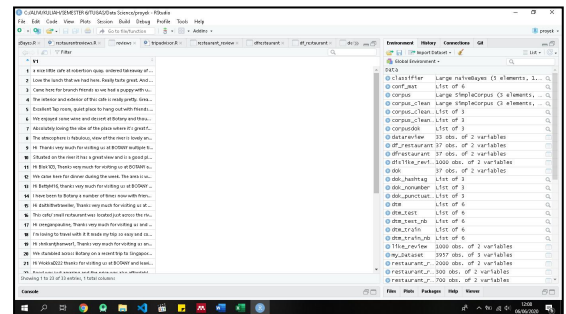
Setelah pengambilan data nama restoran dan juga link untuk review setiap nama restoran selanjutnya adalah pengambilan data *review* dan juga *reviewer* dari setiap restoran, dan berikut adalah tampilan dari hasil data review beserta reviewer yang diambil :



Gambar 4 data reviews dan reviews

## 4. Data review restoran

Terakhir setelah pengambilan data review dan reviewer dari setiap restoran maka dilakukan proses pengambilan data khusus untuk review restoran dan data inilah yang akan menjadi data testing untuk menghasilkan analisis sentiment dan selanjutnya data diolah dengan algoritma naïve bayes



Gambar 5 data reviews restoran

## 3.2 Algoritma Naïve Bayes

### 1. Hasil prediction dan actual Review restoran

Pada tahap ini menampilkan hasil dari perbandingan actual dan prediction untuk Review Restoran. Dengan menampilkan hasil perbandingan antara prediksi dan actual, maka dapat mengetahui data yang berbeda antara prediksi dan actual pada Review restoran dilihat pada Gambar 6.

	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 6 hasil prediksi dan aktual review restoran

Hasil klasifikasi pada penyeleksian Review restoran diklasifikasikan dengan benar pada class 0 ada 111 review. Untuk hasil klasifikasi salah pada class 0 ada 41 review. Untuk hasil klasifikasi benar pada class 1 ada 109 review. Untuk hasil klasifikasi salah pada class 1 ada 39 review.

## 2. Hasil confusion Matrix Review restoran

Pada tahap ini menampilkan hasil confusion matrix dari klasifikasi review restoran yang telah dilakukan. Hasil confusion matrix dari review restoran dapat dilihat pada gambar 7.

	Actual	
Predictions	0	1
0	111	41
1	39	109

Gambar 7 confusion matrix

Pada gambar 7, didapat hasil confusion matrix dari review restoran. Jumlah prediksi benar 0 adalah 111, jumlah prediksi salah 0 adalah 41, jumlah benar 1 adaah 109 dan jumlah salah 1 adalah 39.

Untuk menghitung nilai sensitivity yaitu dengan menggunakan rumus:

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

Setelah itu menghitung nilai precision yaitu dengan menggunakan rumus:

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

Setelah itu, menghitung nilai specificity yaitu dengan menggunakan rumus:

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{109}{109 + 41} = 0.7266666666666667$$

## 3. Hasil Akurasi Review restoran

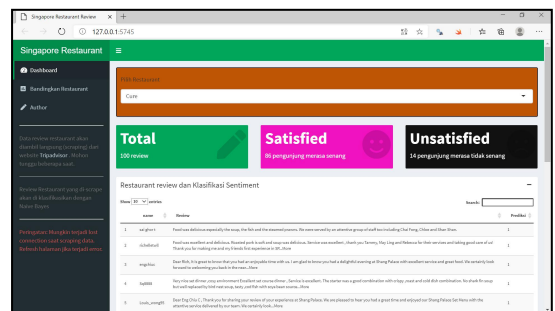
Pada tahap ini menghitung hasil akurasi yang didapatkan dari klasifikasi review restoran.

Untuk menghitung nilai akurasi dapat menggunakan rumus:

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

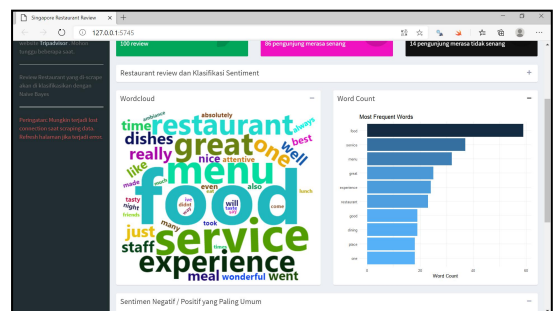
## 3.3 Implementasi dengan Shiny

Pada tahap ini menampilkan output shinyApp, seperti pada gambar berikut ini:



Gambar 8 klasifikasi review restoran

Menampilkan 100 data review yang telah di klasifikasikan dari salah satu restoran siangapura, kemudian hasil klasifikasi 0 dikelompokkan menjadi satiesfied dan 1 menjadi unsatisfied.



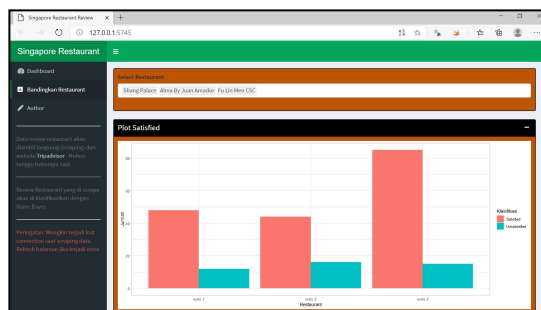
Gambar 9 wordcloud dan word count

Menampilkan wordcloud dari review restoran dan word count(jumlah kemunculan 10 kata yang paling sering muncul).



Gambar 10 sentimen positif/negatif

Menampilkan klasifikasi kata menjadi sentimen positif atau negatif (10 sentimen positif dan 10 sentimen negatif yang paling sering muncul).



Gambar 11 plot satisfied restoran

Menu Banding restoran menampilkan plot satisfied dan unsatisfied dari satu atau lebih restoran singapura.

#### IV. KESIMPULAN

Penelitian proyek akhir untuk menghasilkan analisis sentimen terhadap data restoran dengan menerapkan algoritma Naïve Bayes Classifier yang telah dilakukan menghasilkan kesimpulan bahwa algoritma Naïve Bayes Classifier dapat menentukan klasifikasi dari suatu komentar ke dalam dua kategori yaitu positif yang diimplementasikan dengan kata *Satisfied* atau puas dan juga Unsatisfied atau tidak puas dengan nilai akurasi 73%. Nilai akurasi dari hasil penelitian dipengaruhi oleh proses

pengambilan data pada website, dimana pada satu data reviewer memiliki dua review dengan kelas yang sama. Untuk hasil yang lebih maksimal maka analisis dapat dilakukan dengan data yang baik dan bersih kemudian menggunakan metode lain untuk hasil akurasi yang lebih tinggi.

#### DAFTAR PUSTAKA

- Hanggara, S., Akhriza, T. M., & Husni, M. (2017). Aplikasi Web Untuk Analisis Sentimen Pada Opini Produk Dengan Metode Naive Bayes Classifier. *Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri*, 4(2), A33.1-A33.6.
- Imron, A. (2019). *Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier*.
- Muthia, D. A. (2017). Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 39–45. <https://doi.org/10.1515/HUMOR.2006.009>
- Sinaga, A. (2017). Implementasi Sentiment Analysis untuk Menentukan Tingkat Popularitas Tujuan Wisata. *Prosiding Seminar Nasional Teknologi Dan Rekayasa Informasi Tahun 2017, November*, 24–25. <http://sentrin.fikom.ub.ac.id/event/sentri17/jurnal/pimplementasi-sentiment-analysis-untuk-menentukan-tingkat-popularitas-tujuan-wisata-p/99e80f3>
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi*, 8(1), 958–965.
- Solecha, K. (2019). *Analisa Sentimen Dengan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization Untuk Review Restoran*. 11(1).