

Analisis Sentimen Masyarakat Mengenai Aplikasi ShopeeFood Menggunakan Chi Square dan Metode Naive Bayes Classifier

Muhammad Rizky Ramadhan*, Elvia Budianita, Iwan Iskandar, Muhammad Affandes

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11750115123@students.uin-suska.ac.id, ²elvia.budianita@uin-suska.ac.id, ³iwan.iskandar@uin-suska.ac.id,

⁴affandes@uin-suska.ac.id

Email Penulis Korespondensi: 11750115123@students.uin-suska.ac.id

Abstrak—Pada bulan April 2020, Shopee meluncurkan layanan pesan antar yang dikenal dengan nama ShopeeFood. Layanan ini telah tumbuh menjadi platform layanan besar yang diminati oleh berbagai kelompok usia. Terkait dengan fenomena ShopeeFood, masyarakat memberikan feedback dalam bentuk rating dan opini di berbagai media sosial, salah satunya media Twitter. Langkah yang dilakukan untuk memahami persepsi atau sentimen masyarakat terhadap ShopeeFood, data dikumpulkan melalui media umum dan dianalisis menggunakan metode Naive Bayes Classifier dengan fitur Chi Square. Data tweet diambil melalui Twitter API dengan menggunakan istilah kunci terkait ShopeeFood. Dataset berisi 1050 data yang dibagi menjadi tiga kelas sentimen: positif, negatif, dan netral. Hasil pengujian menggunakan Chi Square pada metode Naive Bayes Classifier dengan rasio pengujian 90:10 dan menghasilkan nilai akurasi sebesar 65,71%, nilai precision 95,00%, dan nilai recall 90,00% dan hasil pengujian menggunakan TF-IDF dengan rasio pengujian 90:10 dan menghasilkan nilai akurasi sebesar 79,05%, nilai precision 95,00%, dan nilai recall 90,00%. Berdasarkan hasil penelitian, dapat disimpulkan bahwa fitur TF-IDF mendapatkan akurasi lebih tinggi dari fitur Chi Square pada metode Naive Bayes Classifier dan berhasil mengklasifikasikan tweet dengan sentimen positif, negatif, dan netral.

Kata Kunci: Chi Square; TF-IDF; Naive Bayes; ShopeeFood; Twitter

Abstract—In April 2020, Shopee launched a delivery service known as ShopeeFood. This service has grown into a large service platform that is in demand by various age groups. Regarding the ShopeeFood phenomenon, the public provides feedback in the form of ratings and opinions on various social media, one of which is Twitter media. The steps taken to understand the perceptions or sentiments of the ShopeeFood community, data are collected through social media and analyzed using the Naive Bayes Classifier method with the Chi feature. Rectangle. Tweet data is retrieved via the Twitter API using key terms related to ShopeeFood. The dataset contains 1050 data divided into three sentiment classes: positive, negative and neutral. The test results use Chi Square on the Naive Bayes Classifier method with a test ratio of 90:10% and produces an accuracy value of 65.71%, a precision value of 95.00%, and the recall value is 90.00% and the test results use TF-IDF with a test ratio of 90:10% and produce an accuracy value of 79.05%, a precision value of 95.00%, and a 90.00% recall value. Based on the research results, it can be concluded that the TF-IDF feature gets higher accuracy than the Chi Square feature in the Naive Bayes Classifier method and is successful in classifying tweets with positive, negative, and neutral sentiments.

Keywords: Chi Square; TF-IDF; Naive Bayes; Shopee Food; Twitter

1. PENDAHULUAN

Pada perkembangan teknologi komputer yang sangat berkembang dengan pesat, hadirilah berbagai inovasi dalam dunia teknologi untuk mendukung kemudahan dalam membantu kehidupan manusia sehari-hari. Mulai dari bidang pembelajaran, kesehatan, teknologi, bisnis, dan bidang lainnya sudah menerapkan inovasi teknologi komputer. Akibat dari inovasi teknologi tersebut, salah satu yang terdampak ialah bidang bisnis, pada bidang bisnis, konsep bisnis berkembang dengan cepat, salah satunya transaksi jual beli dapat dilakukan secara online menggunakan perangkat elektronik (E-Commerce) yang menjadikan proses bertransaksi menjadi lebih praktis dan sederhana.[1]

Salah satu platform *E-Commerce* yang hadir dan besar di Indonesia ialah *Shopee*. *Shopee* merupakan platform *E-Commerce* yang mendukung perkembangan ekonomi digital dan memiliki pasar yang besar di Indonesia. *Shopee* menawarkan berbagai fitur yang membuat *Shopee* menjadikan Indonesia sebagai profit terbesar per tahunnya, pada tahun 2021, *Shopee* diprediksi memiliki keuntungan hingga 67 Triliun, dengan negara penyumbang terbanyak yaitu Indonesia. [2] Salah satu layanan yang ditawarkan PT. *Shopee* Indonesia ialah *food delivery* yang menawarkan jasa pesan antar makanan bernama *Shopee Food*. [3]

Shopee Food launching pada April 2020, namun sejak pertama kali di rilis, *Shopee Food* mampu bertumbuh hingga menjadi platform layanan pesan antar makanan paling banyak digunakan hingga bernilai 22%, pertumbuhan yang cukup cepat dibanding pesaing lainnya yaitu *GoFood* dan *GrabFood*. [2] *Shopee Food* banyak digandrungi oleh berbagai kalangan dari berbagai usia, fitur promo dan discount yang cukup menarik membuat *ShopeeFood* diminati masyarakat Indonesia dan mudah diterima hingga menjadikan *Shopee Food* layanan pesan antar yang cukup diminati. [4]

Berkaitan dengan fenomena *ShopeeFood* masyarakat mencurahkan dan memberikan *feedback* dalam bentuk *rating* dan juga opini tentang *ShopeeFood* pada berbagai sosial media, salah satunya media sosial Twitter. Berdasarkan data laporan dari Global Digital Insights, Indonesia Digital Report 2021 on We Are Social(2021), Indonesia memiliki pengguna aktif media sosial dengan 170 Juta Pengguna dengan total rata-rata menghabiskan waktu 3 jam 14 menit per harinya, salah satu media sosial yang paling aktif digunakan ialah media sosial Twitter dengan 63,6% total pengguna dengan rata-rata waktu pemakaian sebesar 8 Jam per harinya. [5] Melihat fenomena ini, menunjukkan bahwasanya terdapat indikasi peluang sumber data yang sangat besar dan dapat dimanfaatkan untuk menghasilkan sebuah *knowledge* yang

bermanfaat untuk digunakan nantinya, sehingga dalam beberapa kasus, alternatif pengumpulan data melalui media sosial dinilai dapat menarik kesimpulan yang efektif menggantikan survey tradisional. [6]

Berdasarkan hal tersebut diperlukan sebuah algoritma yang dapat memproses sebuah *knowledge* dalam jumlah besar, salah satunya klasifikasi. Algoritma *Naive Bayes* adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema *Bayes*, Penggunaan teorema *Bayes* pada algoritma *Naive Bayes* yaitu dengan mengkombinasikan *prior probilty* dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang memungkinkan. *Naive Bayes* merupakan suatu algoritma yang dapat mengklasifikasi suatu variabel tertentu dengan metode probabilitas dan statistik.[7] Model *independence* ini menghasilkan pemecahan yang terbaik.[8]

Penelitian tentang klasifikasi sentimen telah banyak dilakukan sebelumnya, diantaranya yaitu penelitian yang dilakukan oleh Edwin (2020) pada papernya, Edwin membangun sebuah aplikasi untuk memberikan penilaian (*Assesment*) pada kebijakan pemerintah terkait operasi ojek online di masa pandemi (Covid-19) pada sebuah tweet dengan jumlah 500 data dengan sistem yang digunakan yaitu Naive bayes mendapatkan hasil akurasi 92%.

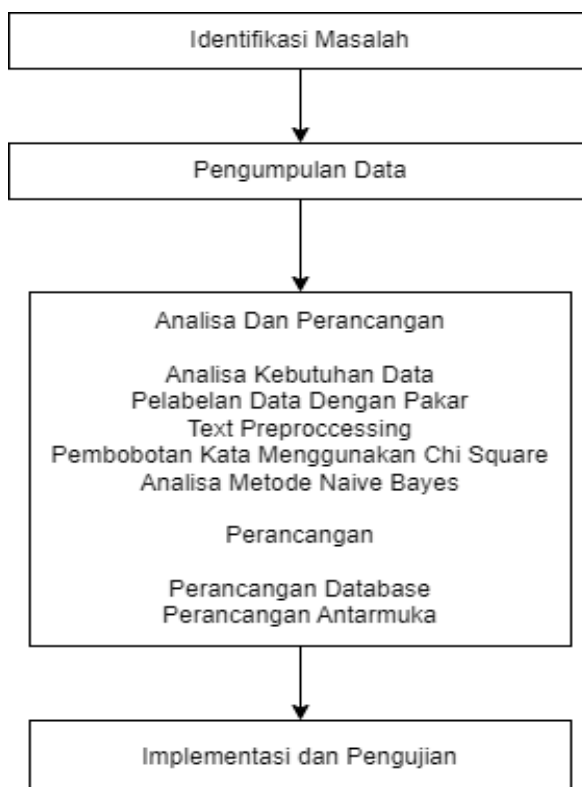
Dengan metode *Naive Bayes*, fitur *Chi Square* merupakan fitur yang bisa digunakan untuk algoritma tersebut. Dalam pengklasifikasian sebuah dokumen, Chi Square adalah salah satu supervised feature selection yang mampu menghilangkan banyak feature tanpa mengurangi tingkat akurasi [9]. Hasil penelitian Alshalabi [10] menunjukkan bahwa penggabungan antara metode klasifikasi Naive Bayes dan pemilihan feature Chi Square dalam klasifikasi teks menghasilkan performa yang sangat baik .[10]

Berdasarkan latar belakang diatas, penulis melakukan penelitian bagaimana menganalisa opini masyarakat terhadap layanan pesan antar *Shopee Food* dengan menerapkan metode algoritma *Naive Bayes Classifier* pada media sosial Twitter. Data yang diambil bersumber dari data tweets yang di ambil menggunakan Twitter API dan dianalisis dibagi kedalam kelas positif negatif dan netral terhadap *ShopeeFood* menggunakan *keyword* dan di proses menggunakan fitur *Chi Square* dan metode *Naive Bayes Classifier*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan kerangka atau susunan kerja dari penelitian secara sistematis agar mencapai tujuan yang diinginkan. Adapun metodologi penelitian dalam tugas akhir ini dipaparkan sebagai berikut:



Gambar 1. Metodologi Penelitian

Pada gambar (1) dijelaskan tahapan-tahapan dalam metologi penelitian yaitu dimulai dari memnentukan permasalahan yang ada atau disebut dengan identifikasi masalah, di lanjutkan dengan pengumpulan data-data, jika sudah mendapatkan data maka akan dilakukan analisa dan perancangan pada data tersebut, dan di akhiri dengan implementasi dan pengujian.

2.2 Perumusan Masalah

Tahap identifikasi masalah ialah menentukan permasalahan yang ada dan melakukan riset terhadap isu ataupun masalah yang melatarbelakangi penelitian. Dalam kasus ini bagaimana menganalisis pro dan kontra masyarakat mengenai *ShopeeFood* menggunakan metode *Naive Bayes Classifier* pada media sosial Twitter.

2.3 Pengumpulan Data

Tahap ini menjelaskan bagaimana cara memperoleh informasi penelitian untuk diproses dan dianalisis. Peneliti melakukan analisa data yang akan diambil dan apa saja yang dibutuhkan terkait. Data dikumpulkan menggunakan data scraper dari tweets pada media sosial Twitter menggunakan Twitter API (*Application Programming Interface*) yang diambil langsung dari media sosial Twitter. Data tweets yang diambil menggunakan tools data scrapping atau data crawling yang merupakan data yang berkaitan dengan penelitian menggunakan kata kunci atau keyword yaitu *ShopeeFood*. [13]

2.4 Analisa dan Perancangan

2.4.1 Analisa

Analisa merupakan tahapan menguraikan detail yang mengandung gambaran penelitian yang akan dilakukan. Berikut adalah tahapan analisa yaitu :

a. Analisa Kebutuhan Data dan Pelabelan Manual Dataset

Analisa kebutuhan data yaitu proses menganalisa data yang akan diproses, data yang diproses ialah data *Twitt* atau data cuitan dari media sosial *Twitter* yang di *Crawling* menggunakan *Twitter API* dan dilabelkan manual kedalam kelas positif, negatif, dan netral lalu di bagi kedalam data latih dan data uji, proses pembagian data latih dan data uji memiliki rasio 90:10% data latih dan data uji. Proses pelabelan data dilakukan dengan cara annotator yang terdiri atas 3 orang.

Berikut adalah contoh data :

Tabel 1. Contoh Data

Data Dummy	Kelas
Mas mas shopeefood baik bgt 😊😊😊	Positif
Gaada yg lebih repot dari membatalkan pesanan shopeefood.	Negatif

b. Text Preprocessing

Text Preprocessing merupakan suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur. *Text Preprocessing* merupakan tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen. Berikut merupakan tahapan dari *Text Preprocessing*: [14]

1. *Cleaning* merupakan proses membersihkan data dari kata-kata yang tidak penting seperti penggunaan karakter, angka, simbol, dan *emoticon*. [15]

Tabel 2. Contoh Proses *Cleaning*

Data sebelum di <i>cleaning</i>	Data sesudah di <i>cleaning</i>
Mas mas shopeefood baik bgt 😊😊😊	Mas mas shopeefood baik bgt
Gaada yg lebih repot dari membatalkan pesanan shopeefood.	Gaada yg lebih repot dari membatalkan pesanan shopeefood

2. *Case Folding* adalah tahapan untuk menyamaratakan penggunaan huruf capital yang diubah menjadi huruf kecil (*lowercase*) semua [16].

Tabel 3. Contoh Proses *Case Folding*

Data sebelum di Case Folding	Data sesudah di <i>Case Folding</i>
Mas mas shopeefood baik bgt	mas mas shopeefood baik bgt
Gaada yg lebih repot dari membatalkan pesanan shopeefood.	gaada yg lebih repot dari membatalkan pesanan shopeefood

3. *Tokenizing* merupakan langkah pemecahan kalimat menjadi kata.

Tabel 4. Contoh Proses *Tokenizing*

Data sebelum di Tokenizing	Data sesudah di Tokenizing
mas mas shopeefood baik bgt	Mas
	Mas
	shopeefood
	Baik

Gaada yg lebih repot dari membatalkan pesanan shopeefood.	Bgt Gaada Yg Lebih Repot Dari membatalkan pesanan shopeefood
---	--

4. Normalisasi merupakan langkah mengubah kata yang memiliki kesalahan eja ataupun mengubah kata-kata tidak baku menjadi kata baku menggunakan kamus normalisasi.

Tabel 5. Contoh Proses Normalisasi

Data sebelum di normalisasi	Data sesudah di normalisasi
Mas	Mas
Mas	Mas
Shopeefood	shopeefood
Baik	Baik
Bgt	banget
Gaada	tidak ada
Yg	Yang
Lebih	Lebih
Repot	Repot
Dari	Dari
membatalkan	membatalkan
Pesanan	pesanan
Shopeefood	shopeefood

5. Lanjutan dari proses normalisasi, proses *Filtering* yaitu mengambil kata-kata yang memiliki unsur penting yang diperoleh dari hasil *tokenizing*.

Tabel 5. Contoh Proses *Filtering*

Data sebelum di Filtering	Data sesudah di Filtering
Mas	Mas
Mas	Mas
Shopeefood	shopeefood
Baik	
Bgt	banget
tidak ada	
Yg	
Lebih	
Repot	Repot
Dari	
membatalkan	membatalkan
Pesanan	pesanan
shopeefood	shopeefood

6. *Stemming* merupakan proses yang dilakukan setelah *filtering*. Pada tahapan ini mengubah kata kembali ke bentuk dasarnya. Proses *stemming* menggunakan algoritma *stemming Enhanced Confix Stripping* (ECS)[17]. Berikut langkah dari *stemming* dengan algoritma *Enhanced Confix Stripping Stemmer* (ECS):

- Mencari kata yang akan distemming pada kamus. Jika kata ditemukan, maka kata tersebut merupakan kata dasar dan algoritma berhenti. Jika tidak lakukan langkah 2.
- Cek *Rule Precedence*. Jika kata memiliki awalan dan akhiran “-lah”, “-an”, “-me-i”, “-di-i”, “-pe-i”, atau “-te-i” langkah selanjutnya yaitu (5, 3, 4, 6). Namun, jika kata tidak memiliki awalan dan akhiran langkah *stemming* berjalan dengan normal.
- Inflexional Particle* (P) (“-lah”, “-kah”, “-tah”, “-pun”) dihilangkan (P) dan kata ganti kepunyaan atau *possessive pronoun* (PP) (“-ku”, “-mu”, “-nya”).
- Hilangkan *Derivational Suffixes* (DS) (“-i”, “-kan”, atau “-an”). Hilangkan *Derivational Prefixes* (DP) (“-di-”, “-ke-”, “-se-”, “-me-”, “-be-”, “-pe-”, “-te-”).
- Mengidentifikasi tipe awalan dan hilangkan. Awalan memiliki dua tipe: Standar: “-di-”, “-ke-”, “-se-”, dapat langsung dihilangkan dari kata. Kompleks: “-me-”, “-be-”, “-pe-”, “-te-” merupakan tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya. Contoh *stemming* dijelaskan pada tabel berikut:

Tabel 7. Contoh Proses Stemming

Data sebelum di stemming	Data sesudah di stemming
mas	Mas
mas	Mas
shopeefood	Shopeefood
banget	Banget
Repot	Repot
Membatalkan	Batal
Pesanan	Pesanan
Shopeefood	Shopeefood

c. Pembobotan Kata Menggunakan *Chi Square*

Term Weighting ialah proses pembobotan kata, yang digunakan untuk memberikan dan menghitung bobot dari setiap kata yang telah diekstrak. Metode yang digunakan pada penelitian ialah *Chi Square* dikarenakan metode *Chi Square* dapat memiliki hasil dan akurasi yang cukup baik. *Chi Square* adalah salah satu supervised feature selection yang mampu menghilangkan banyak feature tanpa mengurangi tingkat akurasi (Sun, Wang, dan Xu, 2009). Hasil penelitian Alshalabi et al (2013) menunjukkan bahwa penggabungan antara metode klasifikasi Naive Bayes dan pemilihan feature *Chi Square* dalam klasifikasi teks menghasilkan performa yang sangat baik (Alshalabi et al, 2013).

Seleksi fitur (feature selection) dilakukan untuk mereduksi fitur-fitur yang tidak relevan dalam proses klasifikasi oleh Naive Bayes. Seleksi fitur *Chi Square* menggunakan teori statistika untuk menguji independensi sebuah term dengan kategorinya. Salah satu tujuan penggunaan seleksi fitur adalah untuk menghilangkan fitur pengganggu dalam klasifikasi. Dalam seleksi fitur *Chi Square* berdasarkan teori statistika, dua peristiwa di antaranya adalah, kemunculan dari fitur dan kemunculan dari kategori, yang kemudian setiap nilai term diurutkan dari yang tertinggi. Uji *Chi Square* dalam statistika diterapkan untuk menguji independensi dari dua peristiwa.

Rumus :

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

O : Nilai Observasi (pengamatan)

E : Nilai Expected (harapan)

$$Df = (b-1) (k-1) \tag{2}$$

D : Jumlah dokumen

b : jumlah baris

k : jumlah kolom

Tahap ini akan melakukan seleksi fitur dengan menggunakan *Chi Square*. Pertama menentukan tabel seleksi *Chi Square* masing-masing fitur dengan tabel berikut:

Tabel 8. Tabel Seleksi Fitur *Chi Square*

	$e_c=1$	$e_c=0$
$e_t=1$	N_{11}	N_{10}
$e_t=0$	N_{01}	N_{00}

Langkah selanjutnya menghitung nilai seleksi fitur *Chi Square*[8] dengan persamaan berikut:

$$\chi^2(D,t,c) = \frac{(N_{00}+N_{11}+N_{10}+N_{01}) \times (N_{00}N_{11} - N_{10}N_{01})}{(N_{00}+N_{11}) \times (N_{00}+N_{11}) \times (N_{00}+N_{11}) \times (N_{00}+N_{11})} \tag{3}$$

Seleksi fitur *Chi Square* digunakan untuk pengamat berkesesuaian (*goodness of fit*) dari kategori dengan *term*. Uji *Chi Square* dalam statistika diterapkan untuk menguji independensi dari dua peristiwa. Sedangkan dalam seleksi fitur berdasarkan teori statistika, dua peristiwa tersebut diantaranya adalah kemunculan dari fitur dan kemunculan dari kategori.[18]

d. Klasifikasi *Naive Bayes Classifier*

Naive Bayes Classifier ialah sebuah metode yang memiliki akar pada teorema bayes. Naive Bayes Classifier memiliki ciri yang utama yaitu terletak pada asumsi yang sangat kuat (naif) akan independensi dari sebuah atributnya. [13]Perbedaan Naive Bayes Classifier dengan teori klasifikasi yang lain ialah pada saat pembangunan hipotesa, pada *Naive Bayes Classifier* hipotesa langsung ditentukan pada suatu frekuensi data latih. Berikut merupakan rumus persamaan *Naive Bayes Classifier*:

$$P(v_j) = \frac{docs_j}{examples} \tag{4}$$

$$P(w_k|v_j) = \frac{1+n_k}{n+|kata|} \tag{5}$$

Keterangan:

- $docs_j$ = kumpulan dokumen yang memiliki nilai target v_j
- examples = jumlah dokumen yang digunakan dalam pelatihan (kumpulan data latihan)
- n = jumlah total kata yang terdapat di dalam data tekstual yang memiliki nilai fungsi target yang sesuai
- n_k = jumlah kemunculan kata w_k pada semua data tekstual yang memiliki nilai fungsi target yang sesuai
- |kata| = jumlah kata yang berbeda yang muncul dalam seluruh data tekstual yang digunakan.

$$V_{MAP} = \underset{v_j \in V}{\text{arg max}} P(V_j) \prod_i P(X_i|v_j) \tag{6}$$

Keterangan:

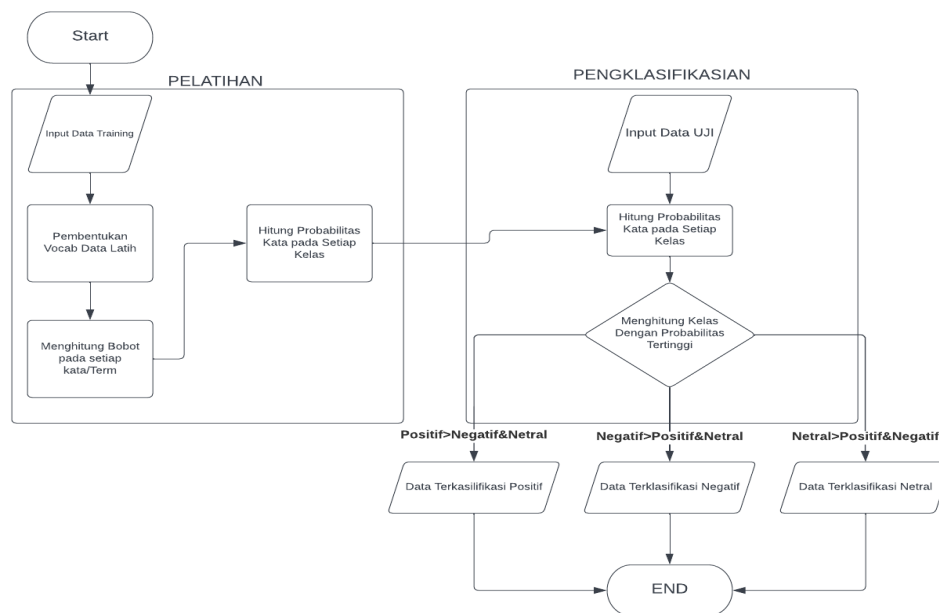
$P(V_j)$: Peluang kemunculan suatu dokumen yang memiliki kategori j

$P(X_i|V_j)$: Peluang kemunculan X_i pada kategori V_j

Berikut adalah tahapan dari Klasifikasi *Naive Bayes Classifier*: [19]

1. Pelatihan
 - a) Melakukan pembentukan *Vocabulary* untuk data latih.
 - b) Menghitung bobot pada setiap kata.
 - c) Menghitung probabilitas setiap kelas.
 - d) Menghitung probabilitas kata pada setiap kelas.
2. Pengklasifikasian
 - a) Menghitung probabilitas dokumen pada masing-masing kelas
 - b) Menentukan kelas dengan nilai probabilitas tertinggi.

Algoritma *Naive Bayes Classifier* (NBC) digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat. Klasifikasi *naive bayes* juga digolongkan ke dalam pembelajaran terawasi (supervised learning) di mana untuk penentuan kelas objek uji didasarkan dari data latihnya dan setiap objek di data latih telah diketahui kelasnya masing-masing seperti pada gambar berikut :



Gambar 2. Flowchart Algoritma *Naive Bayes Classifier*. [20]

2.5 Perancangan

Tahapan ini merupakan tahapan yang bertujuan merancang sistem yang telah dianalisa, dalam tahap perancangan sistem terdapat tiga tahapan, yaitu:

- a. Perancangan tabel atau *database* yang merupakan proses perancangan tabel dan atribut-atribut yang dibutuhkan untuk menyimpan data.
- b. Perancangan struktur menu yang akan dibuat dan digunakan pada sistem.
- c. Tahapan merancang antarmuka atau *interface* untuk memudahkan pengguna atau *user* dalam penggunaan aplikasi yang dibuat.

3. HASIL DAN PEMBAHASAN

Bagian ini memuat langkah-langkah perhitungan menggunakan fitur Chi Square dengan metode Naive Bayes Classifier. Dalam penelitian, terdapat 1050 data uji yang didapatkan dari Twitter API (*Application Programming Interface*).

3.1 Implementasi

Implementasi merupakan tahapan yang dikembangkan berdasarkan hasil analisa dan perancangan yang telah dilakukan sebelumnya. Tahap ini menandakan bahwa aplikasi yang dirancang siap untuk digunakan pada keadaan yang sebenarnya, sehingga akan diketahui apakah sistem akan berjalan dengan tujuan yang diinginkan. Implementasi pada penelitian ini adalah aplikasi yang berbasis web, dimana perangkat pendukung yang dibutuhkan antara lain:

1. Perangkat Keras (*Hardware*)
 - Processor : Intel(R) Core(TM) i5 5200U up to @2.7 GHz
 - Memori (RAM) : 4096 MB DDR
 - Memori (HDD) : 500 GB
2. Perangkat Lunak (*Software*)
 - Sistem Operasi : Microsoft Windows 8.1 Pro 64-bit
 - Data Base : MySQL
 - Web Server : Apache 2.4.39
 - Text Editor : PHP Storm
 - Bahasa Pemrograman : PHP
 - Web Browser : Mozilla Firefox

Kemudian, setelah dilakukan implementasi maka dilakukan tahap pengujian terhadap aplikasi yang telah dirancang. Tahap pengujian dilakukan untuk tujuan mengetahui bagaimana kinerja aplikasi yang telah dicapai. Dalam pengujian kode program akan menggunakan perhitungan tingkat akurasi metode melalui metode pengujian *Confussion Matrix*.

a. Beranda

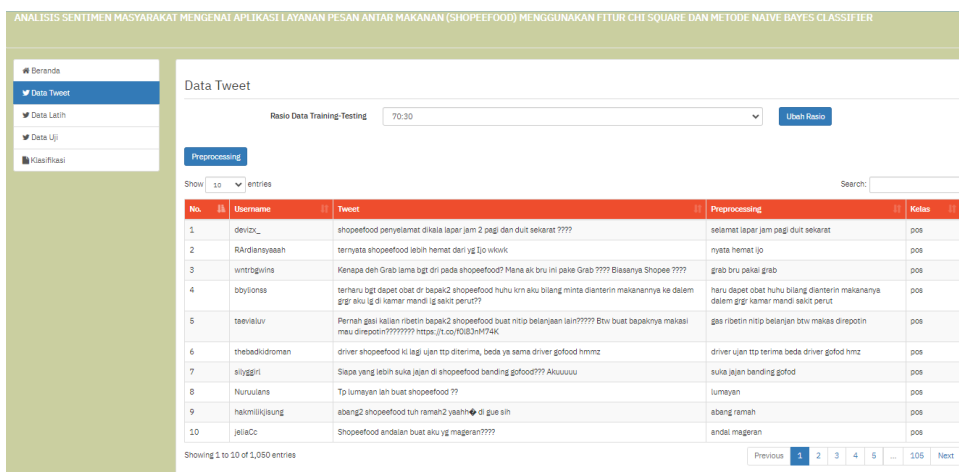
Halaman ini berfungsi untuk menampilkan halaman utama pada sistem, pada halaman ini memiliki beberapa fitur yaitu Beranda, Data tweet, Data latih, Data uji, dan Klasifikasi yang terdapat pada gambar (3).



Gambar 3. Tampilan Beranda

b. Data Tweet

Halaman ini berfungsi untuk menampilkan data-data yang di ambil dari media sosial twitter. Halaman ini terdapat pada gambar (4)



Gambar 4. Data Tweet

c. Hasil Confusion Matrix

Setelah sistem dijalankan, maka mendapatkan hasil akurasi pada Chi Square dan TF-IDF sebagai berikut:

1. Chi Square

Hasil Confusion Matrix menggunakan pembobotan kata Chi Square pada metode Naive Bayes Classifier menggunakan rasio 90:10 yaitu:

Tabel 9. Hasil Confusion Matrix ChiSquare

	Neg	Net	Pos
Neg	34	6	0
Net	5	25	0
Pos	22	3	10

Nilai akurasi menunjukkan ketepatan hasil proses klasifikasi data secara benar. Didapatkan data negatif 34, Netral 25, dan Positif 10. Dapat dikatakan nilai akurasi merupakan perbandingan data yang terklasifikasi benar terhadap keseluruhan data. Selanjutnya dilakukan proses dalam mendapatkan Akurasi setelah menggunakan 3 rasio yaitu 70:30, 80:20, 90:10 adalah sebagai berikut :

Tabel 10. Akurasi ChiSquare

No.	Perbandingan Data Latih dan Data Uji	Akurasi
1	70:30	52,06%
2	80:20	62,38%
3	90:10	65,71%

2. TF-IDF

Hasil Confusion Matrix menggunakan pembobotan kata TF-IDF pada metode Naive Bayes Classifier dengan menggunakan rasio 90:10 adalah sebagai berikut:

Tabel 11. Hasil Confusion Matrix TF-IDF

	Neg	Net	Pos
Neg	38	1	1
Net	10	18	2
Pos	5	3	27

Akurasi yang didapat setelah menggunakan 3 rasio yaitu 70:30, 80:20, 90:10 adalah:

3. TF-IDF pada Naive Bayes

Tabel 12. Akurasi TF-IDF

No.	Perbandingan Data Latih dan Data Uji	Akurasi
1	70:30	68,89%
2	80:20	69,52%
3	90:10	79,05%

Maka didapatkan hasil dari akurasi dengan menggunakan 3 rasio yaitu 70 :30 mendapatkan akurasi 68,89% , dan dengan menggunakan 80:20 mendapatkan akurasi 69,52% dan terakhir dengan rasio 90:10mendapatkan akurasi sebesar 79,05%

4. KESIMPULAN

Berdasarkan pengujian yang telah dilakukan dengan *White Box*, Sistem Analisis Sentimen Masyarakat Mengenai Aplikasi *Shopeefood* Menggunakan *Chi Square* Dan Metode *Naive Bayes Classifier* sangat sesuai. Dengan menggunakan fitur-fitur tersebut, mendapatkan hasil pengujian menggunakan *Chi Square* pada metode *Naive Bayes Classifier* dengan rasio pengujian 90:10% dan menghasilkan nilai akurasi sebesar 65,71%, nilai precision 60,00%, dan nilai recall 81,82% dan hasil pengujian menggunakan TF-IDF dengan rasio pengujian 90:10% dan menghasilkan nilai akurasi sebesar 79,05%, nilai precision %, dan nilai recall %. Berdasarkan hasil penelitian, dapat disimpulkan bahwa fitur TF-IDF mendapatkan akurasi lebih tinggi dari fitur Chi Square pada metode Naive Bayes Classifier dan berhasil mengklasifikasikan tweet dengan sentimen positif, negatif, dan netral.

REFERENCES

- [1] F. S. Mufidah *et al.*, "Analisis Sentimen Masyarakat terhadap Layanan Shopeefood Melalui Media Sosial Twitter dengan Algoritma Naive Bayes Classifier," vol. 7, no. 1, pp. 14–25, 2022, doi: 10.33633/joins.v7i1.5883.
- [2] V.A.R.Barao, R.C.Coata, J.A.Shibli, M.Bertolini, and J.G.S.Souza, "Perbandingan Kinerja Metode Machine Learning Antara Model Naive Bayes Dan K-Nearest Neighbor Terhadap Persepsi," *Braz Dent J.*, vol. 33, no. 1, pp. 1–12, 2022.
- [3] R. Nuryanti, "ShopeeFood, Banjir Diskon Makanan Online Konten ini telah tayang di Kompasiana.com dengan judul

- 'ShopeeFood, Banjir Diskon Makanan Online', Klik untuk baca: <https://www.kompasiana.com/rahmanuryanti1985/60c4e26dd541df78322328d2/shopeefood-banjir-diskon-mak>," 2021, [Online]. Available: <https://www.kompasiana.com/rahmanuryanti1985/60c4e26dd541df78322328d2/shopeefood-banjir-diskon-makanan-online>.
- [4] A. Khomariyah, "Peluang Bisnis Makanan dan Minuman di Shopee Food Bagi UMKM," pp. 1–10, 2021.
- [5] F. Javier, "Layanan Pesan Antar Makanan Paling Banyak Digunakan," 2021, [Online]. Available: <https://data.tempo.co/data/1295/gofood-platform-pesan-antar-makanan-paling-banyak-digunakan>.
- [6] J. HU, W. XU, J. GUO, and W. DENG, "Study on feature selection methods in Chinese text categorization [J]," *Study Commun.*, vol. 3, no. October, p. 014, 2005, [Online]. Available: http://en.cnki.com.cn/Article_en/CJFDTOTAL-GTXY200503014.htm.
- [7] G. P. Kawani, "Implementasi Naive Bayes," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 73–81, 2019, doi: 10.20895/inista.v1i2.73.
- [8] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [9] We Are Social, "Data Reportal : 2021 Indonesia. We Are Social & Hootsuite," 2021. <https://datareportal.com/reports/digital-2021-indonesia> (accessed Jul. 05, 2022).
- [10] E. Edwin, "Aplikasi Assesment Kebijakan Pemerintah Terkait Operasi Ojek Online Di Masa Pandemi (Covid-19) Menggunakan Algoritma Naive Bayes Classifier," *J. Algoritm. Log. dan Komputasi*, vol. 3, no. 2, pp. 299–307, 2021, doi: 10.30813/j-alu.v3i2.2651.
- [11] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.
- [12] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [13] M. Z. Farhan, "Analisis Sentimen Layanan ShopeeFood Pada Twitter Dengan Metode K-Nearest Neighbor , Support Vector Machine , Dan," *J. Ilm. Inform.*, vol. 7, no. 2, pp. 95–106, 2023.
- [14] B. R. Aditya, "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining," *J. INFOTEL - Inform. Telekomun. Elektron.*, vol. 7, no. 2, p. 93, 2015, doi: 10.20895/infotel.v7i2.35.
- [15] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [16] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [17] F. Deviana, "Algoritma Stemming untuk Baso Palembang Berbasis Aturan Tata Bahasa," 2019.
- [18] S. Anisah, A. S. Honggowibowo, and A. Pujiastuti, "Klasifikasi Teks Menggunakan Chi Square Feature Selection Untuk Menentukan Komik Berdasarkan Periode, Materi Dan Fisikdengan Algoritma Naivebayes," *Compiler*, vol. 5, no. 2, pp. 59–66, 2016, doi: 10.28989/compiler.v5i2.171.
- [19] N. Y. Septian, "Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant. 2013*, pp. 1–11, 2009.
- [20] H. Mustofa and A. A. Mahfudh, "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes," *Walisongo J. Inf. Technol.*, vol. 1, no. 1, p. 1, 2019, doi: 10.21580/wjit.2019.1.1.3915.