**Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM)**

Kindi Setiadi1, Nama Penulis Kedua2

1,2 Teknik Informatika – Universitas Mercu Buana

Jl. Meruya Selatan No. 1 Kembangan, Kota Jakarta Barat, 11650

E-mail : 41519120004@student.mercubuana.ac.id1, penulis2@domain2

**A****bstrak**

Banyaknya pengguna internet dan teknologi yang semakin canggih saat ini, kita dapat dengan cepat menemukan apa yang kita inginkan melalui E-Commerce dan dapat digunakan sebagai peluang bisnis dengan membuka toko online. Salah satunya adalah Shopee yang telah merilis layanan ShopeeFood sebagai layanan pesan antar makanan secara online yang dapat memudahkan masyrakat. Promosi merupakan hal yang penting dalam dunia bisnis karena merupakan kunci keberhasilan suatu usaha dalam memasarkan produknya. Dengan kemajuan teknologi yang cepat, media sosial menjadi bagian penting untuk memasarkan agar dapat dikenal. Semakin banyaknya minat masyarakat terhadap layanan tersebut. Dalam hal ini masyarakat banyak memberikan pendapat mereka menggunakan media sosial terutama twitter. Tujuan penelitian ini adalah dapat menganalisis opini masyarakat yang nantinya dapat diklasifikasin oleh metode naïve bayes dan support vector machine, kemudian hasil klasifikasi terdapat 2 kelas negatif dan positif. Metode penelitian yang digunakan adalah pengumpulan data, pembersihan data, dan klasifikasi. Adapun hasil perfoma analisis sentimen menggunakan algoritma naïve bayes dengan tingkat akurasi 90.43%, presisi 99.74%, recall 76.65%, dan f1-score 86.68%. Sedangkan algoritma support vector machine memiliki tingkat akurasi 99.27%, presisi 99.60%, recall 98.60%, dan f1-score 99.10%.

**Kata kunci :** shopeefood, analisis sentimen, naïve bayes, support vector machine

***Abtract***

*Most people the internet and technologies more advanced now, we can quickly find what we want from e-commerce and can be used as online shop. One of them is Shopee which has release ShopeeFood service as service food delivery. Promotion is thing on the bussines world because key to the success of business advertisement. With progress of the technology, social media be an imortant part to advertise the product. In this case most people give a lot of opinion in social media Twitter. Research purposes is can analyze opinion people and can be classified to naïve bayes and support vector machine and then the result has two class negative and positive. The research method used is data collection, data cleansing, and classification. As for the performance results of sentiment analysis using algorithm naïve bayes with level accuracy 90.43%, precission 99.74%, recall 76.65%, dan f1-score 86.68%. While the algorithm support vector machine have level accuracy 99.27%, presisi 99.60%, recall 98.60%, dan f1-score 99.10%..*

***Keywords :*** *shopeefood, sentiment analysis, naïve bayes, support vector machine*

1. **PENDAHULUAN**

Di masa yang sangat berkembang ini, penggunaan internet tidak hanya untuk berkomunikasi tetapi juga untuk berbisnis, memesan transportasi, membeli makanan dan minuman, bekerja, bahkan membeli barang. Keinginan konsumen untuk membuat penilaian yang cepat akan memiliki pengaruh langsung atau tidak langsung terhadap perilaku konsumen saat membeli dan mengkonsumsi barang yang dibeli secara online. Karena banyaknya orang yang menggunakan internet, kita dapat dengan cepat menemukan apa yang kita inginkan melalui *E-Commerce*, yang juga dapat digunakan sebagai peluang bisnis dengan membuka toko online. Salah satu aplikasi *E-Commerce* yang dapat membantu memenuhi kebutuhan manusia, yaitu aplikasi Shopee.

Pada tahun 2017, Shopee melayani hingga 244,8 juta pesanan, dengan pertumbuhan 604,5 juta pesanan pada tahun 2018. Pada tahun 2018, sebanyak 206,9 juta pesanan dilakukan, naik menjadi 98,3 juta pada tahun 2017. Pembeli aktif Shopee berjumlah 49,9 juta pada tahun 2018, naik dari 21,7 juta pada tahun 2017, menunjukkan bahwa penjual aktif telah melonjak [1]. Pada April 2020, Shopee merilis ShopeeFood sebagai layanan *food delivery services*. ShopeeFood, memliki fungsi tidak jauh beda seperti halnya Gofood dan Grabfood, yang merupakan layanan pesan antar makanan dan minuman. Dengan menambahkan fitur yang terdapat di ShopeeFood seperti pengiriman gratis, cash back, promosi, diskon 10%, 15%, 50%, 60%, dan 75% saat menggunakan Shopee Pay, atau diskon 100% untuk pelanggan baru, ShopeeFood telah menjadi sangat populer di kalangan konsumen[2].

Promosi merupakan hal yang penting dalam dunia bisnis karena merupakan kunci keberhasilan suatu usaha dalam memasarkan produknya. Dengan kemajuan teknologi yang cepat, media sosial menjadi bagian penting untuk memasarkan agar dapat dikena l[3]. Media sosial merupakan salah satu wadah atau media untuk berekspresi dan berpendapat atas berbagai tema di era digital saat ini. Media sosial menjadikan seperangkat teknologi komunikasi dan kolaborasi baru yang memungkinkan orang untuk terlibat dalam berbagai bentuk interaksi yang sebelumnya tidak tersedia untuk individu biasa. Salah satu media sosial yang popular saat ini adalah Twitter[4].

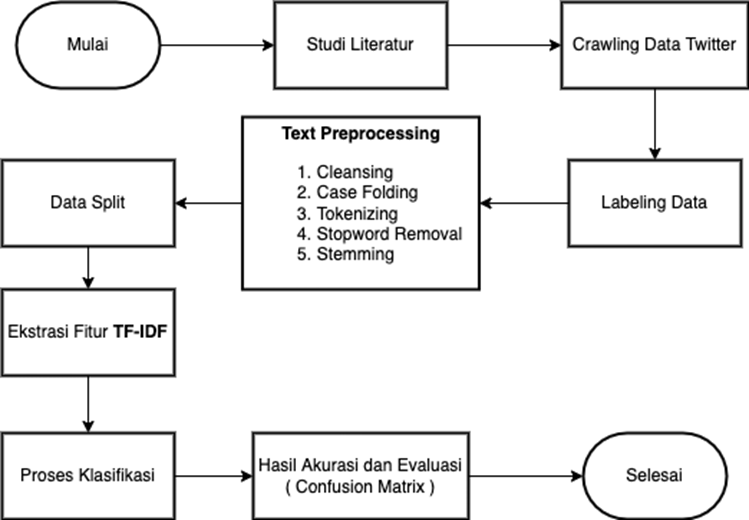
Pada saat ini Twitter semakin populer sebagai alat komunikasi yang ampuh di kalangan pengguna internet. Berdasarkan konferensi resmi pengembang Twitter Chirp 2010, Twitter memiliki 106 juta akun dan 180 juta pengunjung unik bulanan pada April 2010. Jumlah pengguna Twitter diperkirakan meningkat 300.000 per hari[5]. Pentingnya peran media sosial seperti Twitter dalam memposting *tweet*, dapat mengetahui sentimen opini dari pelanggan dengan cara analisis sentimen. Analisis sentimen telah digunakan untuk menentukan reaksi pelanggan terhadap suatu produk, preferensi politik, dan pemikiran tentang film yang baru dirilis, serta proyeksi keuntungan film, dan sebagainya[6].

Berdasarkan penelitian sebelumnya dengan judul “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes”. Hasil penelitian tersebut memiliki pola persepsi negatif pada dataset tweet pemilu 2019 adalah 52%, jauh lebih besar dari pola persepsi positif 18%, dan pandangan netral memiliki nilai 31% lebih tinggi daripada persepsi positif. Akurasi dataset training adalah 81%, sedangkan akurasi dataset testing adalah 76%. Rata-rata nilai presisi sentimen positif sebesar 86,65%, sentimen negatif sebesar 77,15%, dan sentimen netral sebesar 80,95 %, sedangkan nilai rata-rata recall pada sentimen positif sebesar 36,8%, sentimen negatif sebesar 93,2 %, dan sentimen netral sebesar 86,8 %[7].

Kemudian penelitian lainnya dengan judul “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19”. Dalam penelitian ini, melakukan perbandingan hasil analisis sentimen dari vaksin Sinovac dan vaksin merah-putih. Sumber data yang digunakan dari media sosial Twitter. Data yang digunakan sebanyak 845 tweet, dengan menggunakan dua kata kunci, yaitu “vaksinmerahputih” dan “vaksinsinovac”. Setelah itu, data dipisahkan menjadi 253 set pelatihan dan 592 set tes. Algoritma *Support Vector Machine* dan *Naive Bayes* digunakan untuk klasifikasi. Pendekatan Naïve Bayes menghasilkan hasil klasifikasi dengan akurasi rata-rata 85,59%, sedangkan metode SVM menghasilkan hasil dengan akurasi rata-rata 84,41%. Algoritma *Naïve Bayes* mengembalikan 66% sentimen positif dan 34% sentimen negatif untuk istilah "vaccinovac", sedangkan "vaccinmerahputih" mengembalikan 89% sentimen positif dan 11% sentimen negatif. Pendekatan SVM mengembalikan 96% sentimen positif dan 4% sentimen negatif untuk istilah "vaccinovac", sedangkan "vaksinmerahputih" mengembalikan 98% sentimen positif dan 2% negatif[8].

1. **METODOLOGI**

Tahapan penelitian digunakan untuk menunjukkan bagaimana menggambarkan dan menyelesaikan masalah penelitian melalui banyak tahapan. Berikut ini disajikan rancangan atau tahapan - tahapan penelitian serta penjelasan yang disampaikan oleh penulis :



**Gambar 1.** *Flowchart* Tahapan Penelitian

1. **Studi Literatur**

Pada penelitian ini penulis mencari sumber – sumber yang berkaitan *text mining* dan analisis sentimen dengan menggunakan algoritma *naïve bayes* dan *support vector machine*. Literatur yang digunakan sebagai referensi adalah dari jurnal “Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection Algoritma Naive Bayes dan Support Vector Machine”[9]. Selain itu, penulis juga mempelajari buku “Bahasa Pemrograman Python” oleh Abdul Kadir.

1. **Pengumpulan Data / *Crawling Data***

Penulis melakukan pengambilan sumber data dari kumpulan *tweet* menggunakan Twitter API dengan bantuan *library* Tweepy. Pencarian kumpulan data tweet menggunakan kata kunci '("Shopee Food" OR "ShopeeFood") lang:id -has:links', artinya akan mencocokkan frasa persis di isi dalam *tweet* dan data *tweet* hanya Bahasa Indonesia serta tidak mengandung url / link . Serta jumlah data yang diambil berkisar 2000 data *tweet* dari tanggal 02-06-2022 s/d 17-06-2022, kemudian data disimpan dalam format csv.

1. ***Text Preprocessing***

Setelah proses pengambilan data selesai, kemudian pada dataset dilakukan tahapan *text preprocessing* yang merupakan langkah awal dalam pembersihan data sesuai dengan kebutuhannya agar *text mining* dapat mengolahnya lebih lanjut. Algoritma *text mining* sering beroperasi dengan cara yang sama seperti teknik *data mining*. *Text mining* dapat digunakan untuk mengatasi berbagai masalah, termasuk *analysis, classification, clustering,* atau *prediction* dan *information retrieval*[10].

1. *Cleansing*,teknik menghilangkan kata-kata yang tidak perlu dari sebuah teks untuk mengurangi *noise* ﻿seperti Kata, tanda baca, link url, huruf, serta emoji.
2. *Case folding*, Semua karakter huruf atau teks diubah dari huruf kapital menjadi huruf kecil.
3. *Tokenizing*, proses membagi urutan karakter menjadi komponen yang lebih kecil (kata/frasa) yang dikenal sebagai token.
4. *Stopword removal*, membuang kata-kata yang tidak memiliki arti atau tidak relevan, yaitu kata apa yang digunakan untuk merepresentasikan dokumen.
5. *Stemming*, langkah-langkah untuk mengurangi jumlah indeks yang berbeda dalam sumber data sehingga istilah *suffix* atau *prefix* kembali ke bentuk dasarnya.
6. ***Data Split***

Kumpulan data untuk penelitian ini akan dibagi menjadi data latih dan data uji. Pada penelitian ini diterapkan teknik *train test split* menggunakan modul *scikit learn* *library* dengan bahasa pemrograman Python. Jumlah data yang digunakan dalam teknik sebesar 80% data latih dan 20% data uji.

1. **Ekstrasi Fitur TF-IDF**

Setelah melakukan tahapan *text preprocessing* dan *splitting data*, kemudian penulis melakukan tahap selanjutnya yaitu mengg*unakan ﻿Term Frequency - Inverse Document Frequency* (TF-IDF). TF-IDF merupakan proses yang memberikan bobot pada setiap kata untuk menentukan seberapa dekat istilah terhubung ke dokumen[11]. Tujuan dari TF adalah untuk menampilkan frekuensi suatu istilah dalam *tweet*. IDF berusaha untuk menentukan seberapa sering sebuah kata muncul di semua *tweet*. Berikut adalah proses perhitungan pembobotan kata dengan TF-IDF[12].

1. ***Naïve Bayes***
2. ***Support Vector Machine* (SVM)**

**Tabel 1.** Contoh penulisan judul tabel

|  |  |
| --- | --- |
| **Parameter** | **Nilai** |
| Maksimal Perulangan | 10, 50, 100 |
| Rasio Pembelajaran | 0.1 , 0.5, 1 |
| Minimal Eror | 0.1, 0.01, 0.001 |

Sedangkan rumus, dapat dituliskan seperti yang ditunjukkan pada rumus (1) dan (2). Rumus dituliskan rata kiri-kanan dan diberikan nomor rumus di bagian kanannya. Sebaiknya rumus disertai dengan keterangan rumus yang menjelaskan variabel-variabel yang terdapat pada rumus tersebut.

(1)

(2)

Dimana :

ui(ai) : nilai utility dari kriteria i

cout(i)  : nilai kriteria ke-i

cmin : nilai minimal kriteria

cmax : nilai maksimal kriteria

1. **HASIL DAN PEMBAHASAN**

Bagian ini berisi hasil-hasil penelitian dan uraian teoritik yang dapat disajikan dalam bentuk tabel, grafik, gambar, ataupun diagram yang sebaiknya dijelaskan dalam beberapa kalimat. Bagian HASIL DAN PEMBAHASAN dapat terbagi dalam beberapa sub pokok bahasan yang dapat disesuaikan dengan kebutuhan penulisan naskah.

## 3.1 Sub Pokok Bahasan 1

Judul sub pokok bahasan ditulis dengan jenis huruf *Time New Roman* 10pt yang dicetak tebal. Penomoran judul menggunakan angka arab. Setiap huruf awal setiap kata pada judul sub pokok bahasan ditulis dalam huruf besar.

## 3.2 Sub Pokok Bahasan 2

Dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst dst.

1. **PENUTUP**

Bagian ini berisikan kesimpulan berdasarkan hasil-hasil penelitian yang diperoleh, serta kelebihan dan kekurangan dari penelitian yang dilaksanakan. Dapat pula ditambahkan beberapa saran pengembangan untuk penelitian selanjutnya. Bagian ini ditulis dalam bentuk paragraf, tidak dalam bentuk poin-poin.

**DAFTAR PUSTAKA**

1. Z. Hidayat, A. Saefuddin, dan S. Sumartono, “Motivasi, Kebiasaan, dan Keamanan Penggunaan Internet,” *J. ILMU Komun.*, vol. 13, no. 2, 2017.
2. L. Puspitasari dan K. Ishii, “Digital divides and mobile Internet in Indonesia: Impact of smartphones,” *Telemat. Informatics*, vol. 33, no. 2, 2016.
3. R. Kustijono, T. Sunarti, dan H. Budiningarti, “Penggunaan Facebook Sebagai Media Inovatif Dalam Pembelajaran Smp Dan Sma Di Perguruan Muhammadiyah Wiyung Surabaya,” *J. ABDI*, vol. 3, no. 2, 2018.
4. E. Sutinah, G. N. Azima, dan E. F. Imaduddin, “Sistem Informasi Monitoring Akademik Dan Prestasi Siswa Dengan Metode Waterfall,” *J. Inf. Eng. Educ. Technol.*, vol. 2, no. 1, 2018.
5. S. K. Shrivastav, B. Hazela, dan ScheenuRizvi, “Generating UML Diagram Using Natural Language Processing And Use Case Diagram,” *IOSR J. Comput. Eng.*, vol. 20, no. 3, 2018.
6. T. Hidayat dan M. Muttaqin, “Pengujian Sistem Informasi Pendaftaran dan Pembayaran Wisuda Online menggunakan Black Box Testing dengan Metode Equivalence Partitioning dan Boundary Value Analysis,” *J. Tek. Inform. UNIS JUTIS*, vol. 6, no. 1, 2018
7. J. Lach, "SBFS: Steganography based file system," in Proc. of the 2008 1st Int. Conf. on Information Technology, IT 2008, 19-21 May 2008, Gdansk, Poland [Online]. Tersedia: IEEE Xplore, http://www.ieee.org. [Diakses: 10 Sept. 2018].
8. H. A. Nimr, "Defuzzification of the outputs of fuzzy controllers," dipresentasikan pada 5th Int. Conf. on Fuzzy Systems, 1996, Cairo, Egypt. 1996.
9. K. Kimura and A. Lipeles, “Fuzzy controller component,” U. S. Patent 14, 860,040, 14 Dec., 2006.
10. Texas Instruments, “High speed CMOS logic analog multiplexers/demultiplexers,” 74HC4051 datasheet, Nop. 1997 [Direvisi Sept. 2002].
11. Texas Instruments, “High speed CMOS logic analog multiplexers/demultiplexers,” 74HC4051 datasheet, Nop. 1997 [Direvisi Sept. 2002].
12. Nurhayati, S.; and Immanudin, I. , “Penerapan Logika Fuzzy Mamdani untuk Prediksi Pengadaan Peralatan Rumah Tangga Rumah Sakit” *Komputika : Jurnal Sistem Komputer*, Vol.8, No.2, 2019.
13. Saepullah, A.; and Wahono, R.S., “Comparative Analysis of Mamdani, Sugeno and Tsukamoto Method of Fuzzy Inference System for Air Conditioner Energy Saving", *Journal of Intelligent System*. Vol.1, No.2, 2015.
14. Bon, A.T.; and Utami, S.F, “An Analytical Hierarchy Process and Fuzzy Inference System Tsukamoto for Production Planning”, *Prosiding: The Business & Management Review.* Vo.5, No.33,2014.
15. Kotimah, Q.; Mahmudy, W.F.; and Wijayaningrum, V.N., “Optimalization of Fuzzy Tsukamoto Membership Function using Genetic Algorithm to Determine the River Water”, *International Journal of Electrical and Computer Wngineering (IJECE),* Vol.7, No.5, 2017.