

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
NAÏVE BAYES TERHADAP KOMENTAR APLIKASI
TOKOPEDIA**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh
Gelar Sarjana Sistem Informasi*



**PROGRAM STUDI SISTEM INFORMASI
UNIVERSITAS NUSA PUTRA
SUKABUMI
2020**

ABSTRACT

Tokopedia is the online shop most visited by Indonesians, with a total visitor number of 1.2 billion, divided from 863.1 million visitors from the mobile web and 329.8 million visitors from the desktop. Even though it is in the top rank and rating 5 gets the most value, of course not all of them give positive comments, some of them give negative comments, the element of user trust plays an important role in the continuity of an online store, sometimes products are sold or purchased through one of the goods e-commerce sites (the product is) does not match the photo in the advertisement and also sometimes the goods are not what you want.

To research and analyze this, we need a method and analysis to classify user comments into several categories, which in this study are positive and negative categories. This study uses the Naive Bayes method to generate positive and negative sentiments towards comments from users of the Tokopedia application on Playstore.

From the results of tests carried out on 500 user comment data using the TF-IDF process and Term Frequency produces the same value, Testing is based on the accuracy value, Precision and Recall, with an accuracy performance value of 91.20%, with a precision value of 1 While on Recall resulting in a value of 87.68%. AUC value 0.944.

Keyword : sentiment analisys, naive bayes, tokopedia

ABSTRAK

Tokopedia merupakan toko *online* yang paling banyak dikunjungi masyarakat Indonesia yakni dengan total pengunjung mencapai 1,2 miliar yang terbagi dari 863,1 juta pengunjung dari *web mobile* dan 329,8 juta pengunjung dari *desktop*. Meskipun menduduki peringkat teratas dan rating 5 mendapatkan nilai terbanyak, tentu tidak semua memberikan komentar positif, sebagian ada yang memberikan komentar negatif, unsur kepercayaan pengguna memegang peranan penting untuk kelangsungan toko *online*, terkadang produk yang dijual atau dibeli melalui salah satu situs *ecommerce* barang (produknya) tidak sesuai dengan foto di iklan dan juga terkadang barangnya tidak sesuai dengan apa yang di inginkan.

Untuk meneliti dan menganalisis hal tersebut maka diperlukan suatu metode dan analisis untuk mengklasifikasikan komentar pengguna ke dalam beberapa kategori, dimana dalam penelitian ini berupa kategori positif dan negatif. Penelitian ini menggunakan metode *Naive Bayes* untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi Tokopedia di *Playstore*.

Dari hasil pengujian yang dilakukan terhadap 500 data komentar pengguna dengan menggunakan proses TF-IDF dan *Term Frequency* menghasilkan nilai yang sama, Pengujian berdasarkan nilai *accuracy*, *Precision* dan *Recall*, dengan nilai *accuracy performance* sebesar 91,20%, dengan nilai *precision* 1 Sementara pada *Recall* dihasilkan nilai 87,68%. Nilai *AUC* 0,944.

Kata Kunci : analisis sentimen, *naive bayes*, tokopedia

BAB 1

PENDAHULUAN

1.1. LATAR BELAKANG

Perkembangan pemakaian *smartphone* di Indonesia saat ini dirasakan sangat pesat, Jumlah penduduk Indonesia yang mencapai 250 juta jiwa adalah pasar yang besar untuk pesatnya pertumbuhan pengguna *smartphone*. Lembaga riset digital marketing Emarketer memperkirakan pada 2018 jumlah pengguna aktif *smartphone* di Indonesia lebih dari 100 juta orang. Dengan jumlah sebesar itu, Indonesia akan menjadi negara dengan pengguna aktif *smartphone* terbesar keempat di dunia setelah Cina, India, dan Amerika.[1]

Menristekdikti menyebutkan angka pengguna *smartphone* di Indonesia kini mencapai sekitar 25% dari total penduduk atau sekitar 65 juta orang [2], angka prosentase ini menjadi peluang pasar yang menarik sehingga beberapa kalangan di dalam maupun luar negeri kemudian mencoba untuk membuat dan memasarkan aplikasi *mobile* yang bermanfaat untuk kehidupan masyarakat. Ketatnya persaingan di industri e-commerce Indonesia membuat para pelaku bisnisnya harus terus menghadirkan inovasi dan langkah strategis demi bisa memenangkan pasar. Salah satu situs *marketplace* lokal terkemuka Tokopedia meluncurkan aplikasi *mobile* untuk *platform Android*. seluruh fitur Tokopedia bisa diakses oleh penjual dan pembeli melalui perangkat ponsel. Ini merupakan aplikasi *mobile*

Tokopedia yang kedua setelah sebelumnya tahun lalu meluncurkan aplikasi mobile untuk perangkat *BlackBerry* 10. Kini Tokopedia mencoba merambah pasar yang lebih luas dengan meluncurkan aplikasi resmi Tokopedia khusus *Android*. Langkah ini tidak mengherankan karena *platform* buatan *Google* ini merupakan *platform* yang paling pesat pertumbuhan dan penggunaannya saat ini di Indonesia.[3]

Menurut hasil survei situs kupon diskon *CupoNation*. Ditahun 2019, *CupoNation* merilis studi tentang situs belanja *online* yang paling banyak dikunjungi *online shopper* Indonesia sepanjang tahun 2019. Dalam studi ini, terungkap 10 situs belanja *online* terpopuler, baik dari *platform e-commerce* lokal



maupun internasional. Di samping itu, *CupoNation* juga menganalisis sumber pengunjung yang datang dari versi *desktop* maupun *web mobile*.

Tabel 1.1 : 10 situs belanja *online* lokal dan asing yang diurut berdasarkan jumlah trafik terbesar di 2019

Toko Online	Jumlah Pengunjung selama 12 bulan (desktop)	Jumlah Pengunjung selama 12 bulan (web mobile)	Jumlah pengunjung Indonesia selama 12 bulan	% pengunjung dari desktop	% pengunjung dari web mobile
Tokopedia	329,823,447	863,056,416	1,192,879,863	27.65%	72.35%
Shopee	172,275,455	664,805,215	837,080,670	20.58%	79.42%
Bukalapak	153,624,428	669,862,991	823,487,419	18.66%	81.34%
Lazada	67,129,891	378,450,756	445,580,647	15.07%	84.93%
Blibli	141,642,937	211,559,283	353,202,220	40.10%	59.90%
JD ID	24,198,152	81,219,193	105,417,345	22.95%	77.05%
Orami	2,197,378	87,660,992	89,858,371	2.45%	97.55%
Bhinneka	14,017,939	48,911,817	62,929,756	22.28%	77.72%
Sociolla	5,189,789	45,984,181	51,173,970	10.14%	89.86%
Zalora	18,312,615	26,192,069	44,504,684	41.15%	58.85%

Sumber : *CupoNation*

Tokopedia merupakan toko *online* yang paling banyak dikunjungi masyarakat Indonesia yakni dengan total pengunjung mencapai 1,2 miliar yang terbagi dari 863,1 juta pengunjung dari *web mobile* dan 329,8 juta pengunjung dari *desktop*. [4]

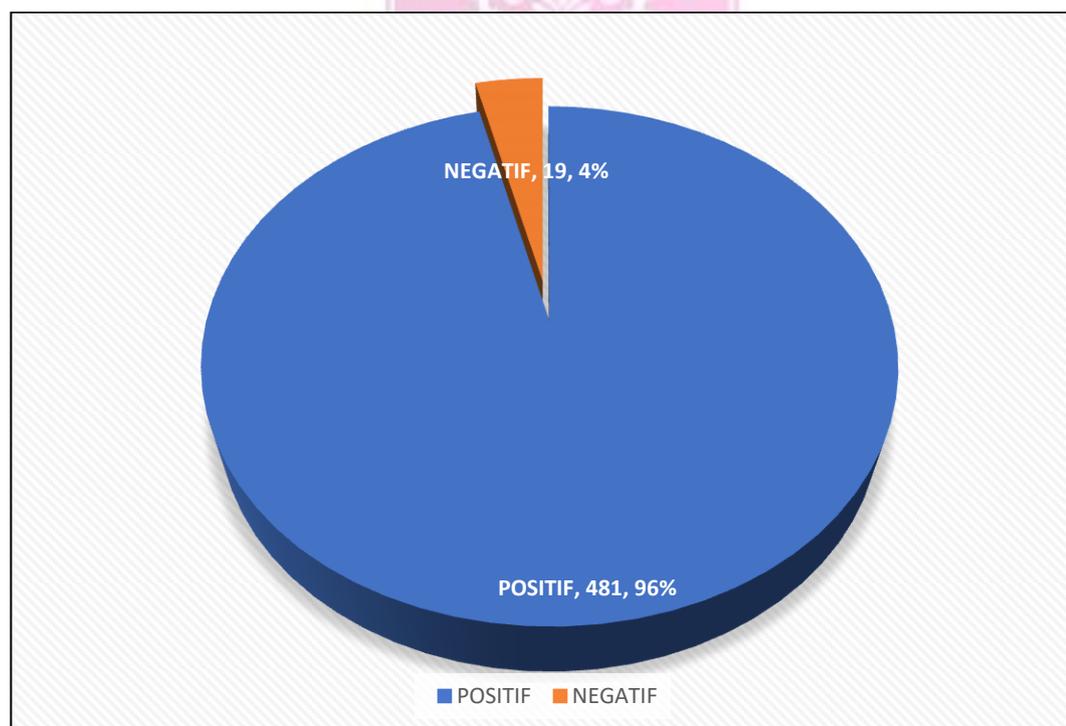
Meskipun menduduki peringkat teratas dan rating 5 mendapatkan nilai terbanyak, tentu tidak semua memberikan komentar positif terhadap aplikasi Tokopedia, mungkin sebagian ada yang memberikan komentar negatif, sebagian lagi memberikan komentar yang netral terhadap aplikasi ini, unsur kepercayaan pengguna atau masyarakat memegang peranan penting untuk kelangsungan toko *online*, terkadang produk yang dijual atau dibeli melalui salah satu situs *ecommerce* barang (produknya) tidak sesuai dengan foto di iklan dan juga

terkadang barangnya tidak sesuai atau bahkan diterima dalam keadaan rusak, tidak sesuai dengan apa yang di inginkan.

Tabel 1.2. Data pemberian rating *user* terhadap Aplikasi tokopedia per tahun 2020

NO	KATEGORI RATING	JUMLAH RATING PENGGUNA
1	Rating Bintang 5	1.941.464
2	Rating Bintang 4	262.360
3	Rating Bintang 3	131.180
4	Rating Bintang 2	78.708
5	Rating Bintang 1	209.888

Sumber : *Google Playstore*



Gambar 1.1. Hasil Analisis Sentimen Positif dan Negatif Terhadap Komentar Di Aplikasi Tokopedia *google Play*

Keterangan : Jumlah komentar yang di *scrap* sebanyak 500 komentar

Pada tabel 1.2 di atas bisa dilihat kategori rating yang diberikan oleh pengguna untuk aplikasi Tokopedia berbanding lurus dengan komentar yang diberikan penggunanya, pada gambar 1.1 terlihat jumlah komentar positif menempati jumlah yang lebih banyak dibandingkan dengan komentar negatif. Untuk meneliti dan menganalisis *objectivitas* hal tersebut maka diperlukan suatu metode dan analisis untuk mengklasifikasikan komentar pengguna kedalam beberapa kategori, dimana dalam penelitian ini berupa kategori positif dan negatif

Dari beberapa penelitian yang pernah dilakukan baik dalam negeri maupun luar negeri, antara lain : Yane Marita Febrianti, Indriati, dan Agus Wahyu Widodo dalam penelitiannya berjudul Analisis Sentimen Pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan *K-Nearest Neighbor (K-NN)* Dengan Perbaikan Kata Menggunakan *Jaro Winkler Distance*. Dengan menggunakan metode *K-NN* Dengan Perbaikan Kata Menggunakan *Jaro Winkler Distance*, pengujian berdasarkan nilai *precision*, *recall*, dan *accuracy* pada masing-masing analisis sentimen tanpa perbaikan kata, maupun dengan perbaikan kata. Hasil pengujian dengan nilai *accuracy* yang baik terdapat pada analisis sentimen dengan perbaikan kata yakni 76 %, dengan nilai *precision* 0,76, dengan nilai *recall* 1[5]. Santoso, Valonia & Virginia, Gloria & Lukito, Yuan. Dalam penelitiannya berjudul Penerapan *Sentiment Analysis* Pada Hasil Evaluasi Dosen Dengan Metode *Support Vector Machine*, penelitian yang dilakukan terhadap hasil evaluasi dosen menghasilkan sentimen analisis positif, negatif dan netral dari hasil angket kuesioner yang diisi oleh mahasiswa menggunakan metode *SVM* dengan hasil penelitian Akurasi tertinggi didapatkan pada dataset dengan $k = 6$ dan akurasi yang dicapai adalah 67,83%. [6]. Sementara Aputra, S. A., Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain. Dalam penelitiannya berjudul Analisis Sentimen *E-Wallet* Pada *Google Play* Menggunakan Algoritma *Naive Bayes* Berbasis *Particle Swarm Optimization*, meneliti tentang ulasan sentimen positif dan negatif pengguna terhadap aplikasi OVO pada *google play/play store*. Algoritma yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier (NB)*, dengan optimasi penggunaan *Feature Selection (FS) Particle Swarm Optimization*. Hasil

dari *cross validation NB* tanpa *FS* adalah 82.30 % untuk *accuracy* dan 0.780 untuk *AUC*. Sedangkan untuk *NB* dengan *FS* adalah 83.60 % untuk *accuracy* dan 0.801 untuk *AUC*. [7].

Penelitian ini memberikan solusi untuk mengetahui tingkat prosentase komentar dan respon pengguna aplikasi tersebut berdasarkan komentar yang diberikan pengguna/konsumen pada aplikasi Tokopedia, dimana dengan prosentase dan respon tersebut akan menjadi masukan demi peningkatan mutu pelayanan, kualitas produk serta kualitas internal perusahaan.

Penelitian ini memberikan manfaat bisa memberikan informasi pada pihak manajemen Tokopedia agar tetap konsisten dalam pelayanan pada konsumen sehingga tingkat kepercayaan pengguna/konsumen tetap terjaga, yang akhirnya dapat meningkatkan keuntungan sesuai target yang dibuat.

Berdasarkan uraian latar belakang diatas, maka peneliti mengambil judul “ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *NAÏVE BAYES* TERHADAP KOMENTAR APLIKASI TOKOPEDIA”

1.2. RUMUSAN MASALAH

Rumusan Masalah dari Penelitian ini adalah sebagai berikut :

1. Bagaimana menerapkan metode Algoritma *Naïve Bayes* dalam sentimen analisis hasil *web scrapping* data komentar dari *google playstore*?
2. Berapa nilai akurasi yang dihasilkan menggunakan metode *Naïve Bayes*?
3. Bagaimana menggambarkan respon pengguna aplikasi Tokopedia ke dalam sebuah klasifikasi berdasarkan komentar pengguna menggunakan aplikasi yang dibuat?

1.3. TUJUAN PENELITIAN

Tujuan Penelitian ini adalah sebagai berikut :

1. Untuk mengetahui penerapan metode *Naïve Bayes* dalam sentimen analisis hasil *web scrapping* data komentar dari *google playstore*
2. Untuk mengetahui Berapa nilai akurasi yang dihasilkan menggunakan metode *Naïve Bayes*.

3. Untuk mengetahui Bagaimana respon pengguna ke dalam sebuah klasifikasi berdasarkan komentar pengguna menggunakan aplikasi yang dibuat

1.4. MANFAAT PENELITIAN

Penelitian ini diharapkan mempunyai manfaat baik secara langsung maupun tidak langsung. Adapun manfaat penelitian ini adalah sebagai berikut :

1.4.1. Manfaat Teoritis :

- a. Memberikan sumbangan pemikiran bagi perkembangan *e-commerce* sesuai dengan tuntutan masyarakat dan sesuai dengan kebutuhan perkembangan jaman terhadap permasalahan-permasalahan *e-commerce* dimasa mendatang.
- b. Sebagai pijakan, bahan referensi dan pengembangan pada penelitian-penelitian selanjutnya yang berhubungan dengan penelitian sentimen analisis

1.4.2. Manfaat Praktis :

- a. Bagi Penulis
Dapat menambah wawasan dan pengalaman langsung tentang cara mengolah data hasil *scraping* komentar *e-commerce* dan mengaplikasikannya dengan menggunakan metode *naive bayes*.
- b. Bagi Masyarakat
Lebih meningkatkan kepercayaan masyarakat terhadap *e-commerce*
- c. Bagi Pemilik *e-commerce*
Bisa menganalisa produknya dan pengembangan aplikasi dimasa depan berdasarkan hasil sentimen analisis tersebut.

1.5. SISTEMATIKA PENULISAN

Sistematika penulisan laporan Tugas Akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan Tugas Akhir ini terdiri dari 5 (Lima) bab, yaitu :

BAB I PENDAHULUAN

Bab ini berisi tentang penjelasan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dari penelitian yang dilakukan, manfaat penelitian dan sistematika penulisan laporan Tugas Akhir.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori-teori mengenai data mining, text mining dan analisis sentimen dengan metode *Naïve Bayes* dan beberapa metode pembandingan lainnya. yang berasal dari jurnal, buku, laporan terdahulu serta studi kepustakaan yang di gunakan sebagai landasan teori dalam pembuatan laporan tugas akhir ini.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang metodologi penelitian dan metodologi dalam menganalisa perhitungan yang digunakan dalam penyusunan Tugas Akhir.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang langkah langkah dalam menganalisis dan proses bagaimana *preprocessing text* dan berisi hasil *Naïve Bayes* serta akurasi.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dan saran dari penelitian Tugas Akhir.

DAFTAR PUSTAKA

Bab ini berisi tentang referensi dan sumber data pada penelitian yang dilakukan

DAFTAR PUSTAKA

- [1]. https://kominfo.go.id/content/detail/6095/indonesia-raksasa-teknologi-digital-asia/0/sorotan_media, [diakses : 5 Juni 2020]
- [2]. <https://www.ristekbrin.go.id/kabar/smartphone-rakyat-indonesia/> [diakses : 5 Juni 2020]
- [3]. <https://dailysocial.id/post/tokopedia-android>. [diakses : 5 Juni 2020]
- [4]. <https://inet.detik.com/business/d-4872404/survei-tokopedia-dan-shopee-ada-kuat-mendominasi> [diakses : 5 Juni 2020]
- [5]. Febrianti, Y., Indriati, I., & Widodo, A. *Analisis Sentimen Pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, p. 3689-3698, feb. 2018. ISSN 2548-964X
- [6]. Santoso, Valonia & Virginia, Gloria & Lukito, Yuan. (2017). *Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine*. *Jurnal Transformatika*. 14. 72. 10.26623/transformatika.v14i2.439
- [7]. aputra, S. A., Didi Rosiyadi, Windu Gata, & Syepry Maulana Husain. (2019). *Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization*. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 377 - 382
- [8]. MUBAROK, Mohamad Syahrul; ADIWIJAYA; ALDHI, Muhammad Dwi. *Aspect-based sentiment analysis to review products using Naïve Bayes*. In: *AIP Conference Proceedings*. AIP Publishing LLC, 2017. p. 020060
- [9]. Gustian, Dudih; Suciati, Indah;, Saepudin, Sudin. *Sistem Pakar Dengan Algoritma Naive Bayes Untuk Prediksi Hasil Produksi Ayam Broiler Plasma (Studi Kasus : Pt.Sekawan Sinar Surya)*, 2019

- [10]. BAYHAQY, Achmad, et al. *Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes*. In: *2018 international conference on orange technologies (ICOT)*. IEEE, 2018. p. 1-6
- [11]. RATNAWATI, Fajar. *Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter*. *INOVTEK Polbeng-Seri Informatika*, 2018, 3.1: 50-59.
- [12]. Kusrini, Luthfi, Emha taufiq, (2009), *Algoritma Data mining*. Yogyakarta:ANDI
- [13]. Hermawati, F.A. (2013), *Data Mining*. Yogyakarta:ANDI
- [14]. Aditya, Bayu Rima. "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining." *Jurnal Infotel 7.2* (2015): 93-100.
- [15]. Nurhuda, Faishol, Sari Widya Sihwi, and Afrizal Doewes. "Analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier." *ITSmart: Jurnal Teknologi dan Informasi 2.2* (2013): 35-42.
- [16]. G, Vinodhini & Chandrasekaran, Dr. (2012). *Sentiment Analysis and Opinion Mining: A Survey*. *Int J Adv Res Comput Sci Technol*. 2.
- [17]. Liu, Bing. "Sentiment analysis and subjectivity." *Handbook of natural language processing 2.2010* (2010): 627-666.
- [18]. Rini, D. C., Farida, Y., & Puspitasari, D. (2016). *Klasifikasi Menggunakan Metode Hybrid Bayesian-Neural Network (Studi Kasus: Identifikasi Virus Komputer)*. *JURNAL MATEMATIKA "MANTIK"*, 01(02), 38-43
- [19]. Lestari, A. R., Perdana, R. S., & Fauzi, M. A. (2017, Desember). *Analisis Sentimen Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan pembobotan Emoji*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1718-1724.
- [20]. Balagatabi, Z. N. (2012) *Comparison Of Decision Tree and Naive Bayes Methods in Classification of Researcher's Cognitive Style in Academic Environment* *Journal of Advance in Computer Research*, (3(2), 23-24.
- [21]. DEY, Lopamudra, et al. *Sentiment analysis of review datasets using naive bayes and k-nn classifier*. *arXiv preprint arXiv:1610.09982*, 2016.

- [22]. ALSAEEDI, Abdullah; KHAN, Mohammad Zubair. A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, 2019, 10.2: 361-374.
- [23]. J. Prager, "Open-Domain Question–Answering," Found. Trends® Inf. Retr., vol. 1, no. 2, pp. 91–231, 2006

