

## ANALISIS SENTIMEN MASYARAKAT TERHADAP THRIFTING PADA MEDIA SOSIAL TWITTER(X.COM) MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Dafa Akbar Firmansyah<sup>1</sup>, Devi Damayanti<sup>2</sup>.

<sup>1</sup>Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: <sup>1</sup>dafaakbarfirmansyah@gmail.com, <sup>2</sup>dosen02390@unpam.ac.id

### Abstrak

Saat ini fenomena thrifting bukanlah hal yang baru dan menjadi populer dikalangan anak muda. Fashion saat ini menjadi kebutuhan yang perkembangannya sangat cepat dengan munculnya berbagai trend fashion baru dikalangan remaja dalam waktu singkat. Salah satu trend adalah fashion thrift. Thrifting adalah aktivitas berbelanja barang bekas dengan tujuan menghemat biaya. Namun, baru baru ini muncul isu tentang larangan thrifting di Indonesia, yang memicu kontroversi publik. Larangan thrifting di Indonesia didasarkan pada Undang Undang Republik Indonesia Nomor 7 Tahun 2014 tentang Perdagangan. Pasal tersebut menyatakan bahwa "setiap importir wajib mengimpor barangnya ke keadaan baru". Dari pemaparan tersebut ada pihak yang setuju ada pula yang tidak. Berdasarkan banyaknya opini masyarakat di Twitter tentang thrifting, kita bisa menggunakan analisis sentimen untuk meninjau opini mana yang lebih dominan di masyarakat. Pada penelitian kali ini penulis ingin melakukan analisis sentimen terhadap topik thrifting menggunakan metode k-nearest neighbor. Hasil klasifikasi menggunakan metode K-Nearest Neighbor mampu mengklasifikasikan respon dari pengguna twitter dan dapat digunakan pemerintah sebagai bahan evaluasi dan penilaian terhadap thrifting. Hasil pengujian metode K-Nearest Neighbor menggunakan confusion matrix dengan data sebanyak 252 mendapatkan tingkat akurasi sebesar 76% dengan nilai k=3.

Kata kunci: Analisis Sentimen, *Text Mining*, *K-Nearest Neighbor*, *Thrifting*, *Twitter*

### Abstract

Currently, the thrifting phenomenon is nothing new and has become popular among young people. Fashion is currently a need that is developing very quickly with the emergence of various new fashion trends among teenagers in a short time. One trend is thrift fashion. Thrifting is the activity of shopping for used goods with the aim of saving costs. However, recently the issue of banning thrifting in Indonesia has emerged, which has sparked public controversy. The ban on thrifting in Indonesia is based on Law of the Republic of Indonesia Number 7 of 2014 concerning Trade. The article states that "every importer is obliged to import his goods to a new condition". From this explanation, there are parties who agree and there are those who don't. Based on the number of public opinions on Twitter about thrifting, we can use sentiment analysis to review which opinions are more dominant in society. In this research, the author wants to conduct sentiment analysis on the topic of thrifting using the k-nearest neighbor method. The classification results using the K-Nearest Neighbor method are able to classify responses from Twitter users and can be used by the government as evaluation and assessment material for thrifting. The results of testing the K-Nearest Neighbor method using a confusion matrix with 252 data obtained an accuracy level of 76% with a value of k=3. Keywords: *Sentiment Analysis*, *Text Mining*, *K-Nearest Neighbor*, *Thrifting*

Keywords: *Sentiment Analysis*, *Text Mining*, *K-Nearest Neighbor*, *Thrifting*, *Twitter*.

## 1. PENDAHULUAN

Fenomena *thrifting* atau gaya hidup hemat dalam berbelanja semakin populer di kalangan masyarakat saat ini. Hal ini dipengaruhi oleh beberapa faktor, seperti kondisi ekonomi yang sulit, adanya pandemi Covid-19 yang membatasi aktifitas dan mempengaruhi pengeluaran, serta semakin mudahnya akses ke informasi dan teknologi. Media sosial Twitter merupakan salah satu media sosial yang banyak digunakan untuk berbagi informasi dan pengalaman mengenai *thrifting*, seperti cara membeli barang dengan memanfaatkan diskon, promo, menggunakan kembali barang bekas.

Dalam konteks pemasaran, fenomena *thrifting* ini bisa menjadi peluang atau tantangan bagi perusahaan atau bisnis. Di sisi lain, perusahaan dapat memanfaatkan tren ini untuk mengembangkan strategi pemasaran yang lebih efektif, seperti dengan menawarkan produk atau layanan yang hemat, atau menawarkan promo dan diskon yang menarik. Di sisi lain, perusahaan juga perlu memahami persepsi masyarakat terhadap *thrifting*, agar tidak terjadi kesalahan strategi pemasaran yang dapat merugikan perusahaan.

Namun, baru baru ini muncul isu tentang larangan *thrifting* di Indonesia, yang memicu kontroversi publik. Larangan *thrifting* di Indonesia didasarkan pada Undang-Undang Republik Indonesia Nomor 7 Tahun 2014 tentang Perdagangan. Pasal tersebut menyatakan bahwa “setiap importir wajib mengimpor barangnya ke keadaan baru”. Di sisi lain, impor barang bekas hanya diperbolehkan dengan syarat-syarat tertentu yang ditetapkan oleh menteri.

Menanggapi hal itu banyak opini berkembang di masyarakat tentang *thrifting*. Ada yang setuju dan ada juga yang tidak setuju. Berdasarkan banyaknya opini masyarakat di Twitter tentang *thrifting*, kita bisa menggunakan analisis sentimen untuk memperkirakan opini mana yang lebih dominan di masyarakat[1].

Dari uraian di atas, peneliti menganalisis mengenai *thrifting*, data yang telah didapatkan kemudian diolah dan diklasifikasikan dengan metode *K-Nearest Neighbor*.

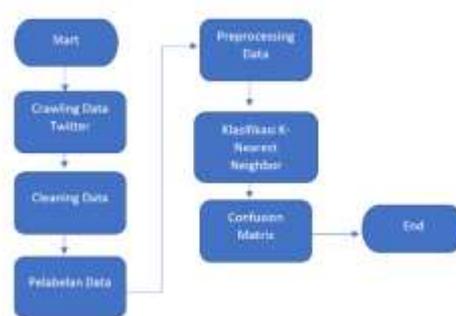
Metode *K-Nearest Neighbor* adalah suatu pendekatan untuk mengelompokkan objek baru dengan mempertimbangkan (K) tetangga terdekatnya. Ini termasuk dalam kategori algoritma *supervised learning*, di mana hasil klasifikasi dari *instance query* yang baru ditentukan oleh mayoritas kategori tetangga terdekat. Klasifikasi hasilnya didasarkan pada kategori yang paling banyak muncul di antara tetangga terdekat, dan meskipun metode ini lebih efektif ketika digunakan pada set data training yang besar untuk menghasilkan prediksi yang lebih akurat, penting untuk menentukan nilai optimal untuk K, yang menunjukkan jumlah tetangga terdekat. Namun, perlu dicatat bahwa metode *K-nearest neighbor* memerlukan komputasi yang cukup tinggi karena perhitungan jarak harus dilakukan untuk setiap instance query bersamaan dengan seluruh instan dari sampel pelatihan[2].

*K-Nearest Neighbor* dapat membantu mengklasifikasikan kondisi objek yang memiliki variabel K berdasarkan data uji pada kondisi jarak-kedekatan Objek data training serta peraturan untuk menentukan nilai variabel K tidak lebih dari data training ditentukan dan nilai K harus ganjil atau tidak kosong[3].

Hasil penelitian ini nantinya dapat bermanfaat bagi rekan rekan mahasiswa, masyarakat dan pemerintah untuk mengetahui banyaknya sentimen positif dan negatif, sehingga nantinya diusahakan untuk evaluasi lebih lanjut[4].

## 2. METODOLOGI PENELITIAN

Metodologi penelitian yang dilakukan pada penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Metode Penelitian

### 2.1 Crawling Data

Data dikumpulkan pertama-tama dengan langkah mendaftar akun Twitter, diikuti dengan pendaftaran sebagai pengembang aplikasi melalui situs `developer.twitter.com`. Langkah ini diperlukan untuk memperoleh akses API tweet yang akan digunakan untuk menarik data tweet pengguna. Proses penarikan data tweet memerlukan kata kunci "*thrifting*" dengan persyaratan bahwa tweet harus dalam bahasa Indonesia. Pengumpulan data dimulai pada rentang tanggal 6-15 Maret 2023 dan 15-24 Juni 2023.

### 2.2 Cleaning Data

Tahap cleaning bertujuan untuk menghapus simbol, tanda baca, angka, url, mention, emoticon, username sehingga menghasilkan kalimat yang akan diproses ke tahap selanjutnya[5].

### 2.3 Pelabelan Data

Sebelum melakukan tahapan preprocessing data, data yang telah dicleaning harus diberikan label untuk menentukan sentimen dari setiap data. Label sentimen setiap data berupa positif dan negatif

## 2.4 Preprocessing Data

Preprocessing Data ini bertujuan untuk membentuk dan membersihkan data sebelum melakukan tahap klasifikasi[6]. Berikut adalah langkah-langkah dalam proses preprocessing:

- a) Case Folding  
Case Folding digunakan untuk mengubah huruf kapital menjadi huruf kecil (lowercase) pada seluruh data dalam dokumen[7].
- b) Tokenize  
Dalam langkah ini, terjadi pemisahan kalimat menjadi kata-kata individu dan dilakukan pemeriksaan kata dari awal hingga akhir karakter[8].
- c) Stopword Removal  
Proses stopwords removal adalah langkah untuk mengambil kata-kata kunci dari hasil tokenisasi. Hal ini dapat dilakukan dengan menggunakan algoritma stoplist untuk menghapus kata-kata yang dianggap kurang penting atau dengan menggunakan wordlist untuk menyimpan kata-kata yang dianggap penting. Stopword adalah kata-kata yang kurang deskriptif dan dapat dihilangkan. Contohnya, "yang", "dan", "di", dan sebagainya[9].
- d) Stemming  
Stemming merupakan proses mencari akar kata dari setiap kata hasil penyaringan dengan menghapus imbuhan baik di awal maupun di akhir kata[10].

## 2.5 Klasifikasi K-Nearest Neighbor

Model klasifikasi K-nearest neighbors adalah suatu pendekatan supervised learning yang digunakan untuk mengklasifikasikan objek berdasarkan variabel dan data pelatihan. Dalam melakukan klasifikasi, diperlukan data uji untuk menghitung jaraknya dengan objek data pelatihan yang paling dekat. Proses perhitungan ini menggunakan euclidean distance untuk mengukur jarak antara data baru dan data pelatihan, dengan menggunakan rumus perhitungan euclidean distance[11].

$$d(x,y) = \sqrt{\sum_{i=1}^n (x - y)^2} \quad (1)$$

- x : dokumen testing atau uji  
y : dokumen training atau latih

## 2.6 Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi klasifikasi dalam metode K-Nearest Neighbors (K-NN). Penggunaan confusion matrix bertujuan untuk memahami seberapa akurat metode yang sedang dievaluasi. Rumus yang digunakan untuk menghitung confusion matrix guna menentukan akurasi, presisi, dan recall adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (3)$$

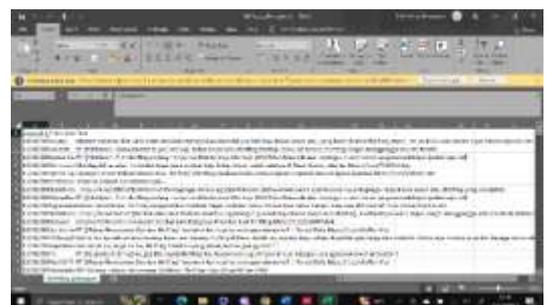
$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

## 3. HASIL DAN PEMBAHASAN

Pada bagian ini berisikan hasil dan pembahasan mengenai pengumpulan data, preprocessing, Analisis Sentimen dan evaluasi.

### 3.1. Crawling Data

Crawling data diambil dari twitter dengan kata kunci "thrifting". Crawling data yang diambil dalam bentuk teks yang merupakan tweet dari pengguna twitter. Data yang diambil pada tanggal 6-15 Maret 2023 dan 15-24 juni 2023 berjumlah 1017 tweet. Dalam crawling data, penelitian ini menggunakan aplikasi RapidMiner yang disimpan dalam bentuk file csv. Peneliti menggunakan operator "Search Twitter" yang disambungkan pada akun twitter tersebut melalui koneksi twitter di rapidminer. Berikut hasil crawling data twitter



Gambar 2 Hasil Crawling Data

### 3.2. Cleaning Data

Cleaning adalah proses untuk membersihkan data tweet yang pada data awalnya memiliki banyak symbol seperti (@!#\$%^&\*:'), kemudian juga menghapus nama akun, dan url sehingga menghasilkan data tweet yang lebih bersih.

**Tabel 1.** Cleaning

Sebelum	Sesudah
@AryaBara @ffikriawan Digaji negara buat thrifting pakaian yang udah dilarang negara ?	Digaji negara buat thrifting pakaian yang udah dilarang negara

Setelah melakukan pembersihan data maka tweet tersisa 664 tweet yang nantinya akan diberi label secara manual.

### 3.3. Pelabelan Data

Pelabelan merupakan tahap menentukan sentimen pada data menjadi positif dan negatif. Pelabelan dilakukan dengan secara manual. Berikut hasil pelabelan manual.



**Gambar 3** Pelabelan

Dari 664 data yang diperiksa, hanya 252 di antaranya memenuhi kriteria dan dapat diberi label sentimen, sementara sisanya tidak relevan.

### 3.4. Preprocessing Data

Text processing merupakan suatu proses yang dilakukan untuk menyeleksi data teks dengan maksud membuatnya lebih terstruktur, sehingga data tersebut dapat disiapkan untuk analisis.

#### 3.4.1 Case Folding

Tahap ini dimaksudkan untuk mengkonversi semua karakter dalam teks menjadi huruf kecil (lowercase), sehingga teks tersebut dapat dianggap memiliki kasus yang seragam dalam proses pemrosesan teks.

**Tabel 2** Case Folding

Sebelum	Sesudah
Digaji negara buat thrifting pakaian yang udah dilarang negara	digaji negara buat thrifting pakaian yang udah dilarang negara

#### 3.4.2 Tokenize

Tokenize adalah langkah untuk membagi atau memotong teks menjadi unit-unit yang lebih kecil yang disebut "token". Token ini merupakan unit dasar dalam analisis teks, umumnya berupa kata atau frasa pendek.

**Tabel 3** Tokenize

Sebelum	Sesudah
digaji negara buat thrifting pakaian yang udah dilarang negara	['digaji','negara','buat', 'thrifting','pakaian', 'yang','udah','dilarang', 'negara']

#### 3.4.3 Stopword Removal

Langkah ini dimaksudkan untuk menghapus term atau kata dasar yang tidak memiliki makna atau tidak relevan. Dalam penelitian ini, penghapusan stopwords menggunakan kamus yang berisi kumpulan stopwords dalam bahasa Indonesia.

**Tabel 4** Stopword Removal

Sebelum	Sesudah
digaji negara buat thrifting pakaian yang udah dilarang negara	['digaji','negara', 'thrifting','pakaian', 'udah','dilarang','negara']

#### 3.4.4 Stemming

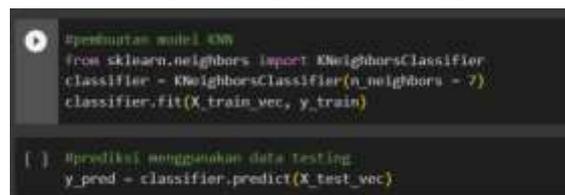
Stemming merupakan langkah untuk menemukan akar kata dari setiap kata yang telah difilter dengan menghilangkan imbuhan, baik itu di depan maupun di belakang kata.

**Tabel 5** Stemming

Sebelum	Sesudah
digaji negara thrifting pakaian udah dilarang negara	['gaji', 'negara', 'thrifting', 'pakai', 'udah', 'larang', 'negara']

### 3.5. Klasifikasi K-Nearest Neighbor

Pada tahapan kali ini akan dilakukan klasifikasi terhadap sentimen dari Twitter mengenai Thrifting dengan menggunakan algoritma K-Nearest Neighbor (KNN). Klasifikasi algoritma K-Nearest Neighbor yang dilakukan pada analisis ini menggunakan rasio 70% data training dan 30% data testing. Pada tahap ini dilakukan uji variasi nilai k dengan menggunakan Python. Angka yang digunakan berupa angka ganjil dari 3 hingga 9.



**Gambar 4** Klasifikasi K-Nearest Neighbor

### 3.6. Confusion Matrix

Berikut adalah tabel hasil confusion matrix dari Klasifikasi Metode K-Nearest Neighbor (KNN)

**Tabel 6** Hasil Klasifikasi

Nilai K	Akurasi	Presisi	Recall
3	0.76	0.62	0.25
5	0.75	0.67	0.10
7	0.75	0.67	0.10
9	0.74	0.50	0.05

## 4. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan, dapat disimpulkan bahwa Hasil sentimen yang didapatkan dari tanggapan masyarakat terhadap thrifting didapatkan sebanyak 177 sentimen positif dan 75 sentimen negatif. Berdasarkan hasil pengujian sentimen positif lebih banyak ketimbang sentimen negatif yang berarti thrifting menjadi kegiatan positif berdasarkan tanggapan masyarakat. Berdasarkan hasil yang diperoleh analisis sentimen dilakukan dengan menerapkan Algoritma K-Nearest Neighbor memiliki nilai akurasi tertinggi sebesar 76% dengan nilai rasio perbandingan sebesar 70:30 dan nilai parameter  $k=3$ .

## 5. REFERENCES

- [1] A. Kurniawan and S. Waluyo, "Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter," *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 455–461, 2022.
- [2] S. Saepudin, M. Muslih, and Sihabudin, "Pemilihan Jurusan Dengan Metode K-Nearest Neighbor Untuk Calon Siswa Baru," *J. Rekayasa Teknol. Nusa Putra*, vol. 5, no. 2, pp. 15–19, 2019.
- [3] A. Halimi, Kusriani, and M. R. Arief, "ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP PEMBELAJARAN ONLINE DARI DI MEDIA SOSIAL TWITTER MENGGUNAKAN LEXICON DAN K-NEAREST NEIGHBOR," vol. 2, no. 1, 2021, [Online]. Available: <https://ejournal.unuja.ac.id/index.php/core>
- [4] N. Ferdiansyah and A. Solichin, "Analisis Sentimen Terhadap Pembelajaran Dosen Berdasarkan Data Kritik Saran Mahasiswa Menggunakan Metode Naive Bayes," *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 19, no. 2, pp. 104–111, 2022.
- [5] F. R. Irawan, A. Jazuli, and T. Khotimah, "ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEARSET NEIGHBORS," *J. Inform. dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 1, 2022, doi: 10.33387/jiko.
- [6] W. Adzhan and D. Yusup, "Implementasi Metode K-Nearest Neighbor Dalam Menganalisis Sentimen Terhadap Penyedia Jasa Layanan Internet First Media," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 594–602, 2022, doi: 10.5281/zenodo.7068128.
- [7] M. Furqan, Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia Sentiment Analysis using K-Nearest Neighbor towards the New Normal During the Covid-19 Period in Indonesia," 2022. [Online]. Available: [www.tripadvisor.com](http://www.tripadvisor.com)
- [8] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," vol. 6, no. 1, 2021, doi: 10.30743/infotekjar.v6i1.3893.
- [9] Z. Ulfah Siregar, R. R. A. Siregar, and R. Arianto, "KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR," vol. 8, no. 1, 2019.
- [10] P. P. O. Mahawardana, I. A. P. F. Imawati, and I. W. Dika, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *J. Manaj. dan Teknol. Inf.*, vol. 12, no. 2, pp. 50–56, 2022, [Online]. Available: <https://ojs.mahadewa.ac.id/index.php/jmti/article/view/2111>
- [11] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.