



Implementasi Metode K-Nearest Neighbor Dalam Menganalisis Sentimen Terhadap Penyedia Jasa Layanan Internet First Media

Wafiuddin Adzhan¹, Hannie², Dadang Yusup³

^{1,2,3}Universitas Singaperbangsa Karawang

Email: wafi.adzhan18137@student.unsika.ac.id

Abstract

Received: 10 Agustus 2022

Revised: 13 Agustus 2022

Accepted: 18 Agustus 2022

At this time the use of the internet is growing rapidly and the internet plays an important role in various aspects of human life. With the increasing number of Internet users like that, many internet service providers are competing to use their services. In addition, the internet can also be used to search for information through social media, one of which is social media Twitter. Twitter is often used to express emotions about something good or bad. Tweets published by users via Twitter contain an opinion about an object, the object can be of various kinds, such as one of them is about a product or service. One of the internet services that is familiar to various groups of people is First Media, which is often given feedback by its users on the quality of the internet service provider. These responses are often given through various kinds of social media, one of which is Twitter and from these responses a method, namely sentiment analysis, will be carried out. This study discusses reviews in the form of tweets related to first media with the K-Nearest Neighbor algorithm using Rapid Miner tools. Data taken as many as 500 tweets related to the first media. Then do the deletion of ambiguous data so that it only produces 300 tweets. The methodology used is Knowledge Discovery in Databases, namely Data Selection, Data Pre-processing, Transformation, Data Mining, Evaluation. Data labeling is carried out by experts who are divided into positive and negative sentiments. The results of the modeling using K-Fold Cross Validation, and testing the value of k variations with odd numbers ranging from 3 to 25 to get the best accuracy. The highest accuracy was obtained with an accuracy of 84% with a value of k=3. Good classification results can be seen by how much accuracy is obtained.

Keywords: *Twitter Sentiment Analysis, Knowledge Discovery in Databases, K-Nearest Neighbor, First Media, Rapid Miner*

(*) Corresponding Author: wafi.adzhan18137@student.unsika.ac.id

How to Cite: Adzhan, W., Hannie, H., & Yusup, D. (2022). Implementasi Metode K-Nearest Neighbor Dalam Menganalisis Sentimen Terhadap Penyedia Jasa Layanan Internet First Media. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 594-602. <https://doi.org/10.5281/zenodo.7068128>

PENDAHULUAN

Pada saat ini penggunaan internet berkembang pesat dan internet berperan penting dalam berbagai segi kehidupan manusia. Dengan jumlah pengguna Internet yang semakin banyak seperti itu maka banyak penyedia layanan internet yang berlomba lomba untuk menggunakan jasanya. Salah satu penyedia layanan Internet atau penyedia jasa Internet Service Provider (ISP) adalah First Media. Dalam mencari informasi yang akurat pengguna internet memanfaatkan media sosial salah satunya seperti media sosial twitter. Pada media sosial Twitter sering digunakan untuk mengungkapkan emosi mengenai suatu hal yang baik maupun buruk. Tweet yang dipublikasikan pengguna melalui Twitter mengandung sebuah opini mengenai sebuah objek, objek tersebut dapat berbagai macam seperti salah

satunya adalah tentang suatu produk atau layanan. Hal ini digunakan perusahaan dalam menyebarkan dan mencari informasi melalui Twitter untuk mengetahui respon dari masyarakat terhadap perusahaan tersebut (Fadhilah Dwi Ananda, Yoga Pristyanto. 2021).

First Media sendiri merupakan salah satu penyedia jasa layanan Internet Service Provider (ISP), TV kabel, dan komunikasi di Indonesia. Dengan pengguna yang tersebar luas di seluruh Indonesia maka banyak sekali respon masyarakat mengenai layanan Provider First Media. Beberapa respon masyarakat juga mengandung sebuah opini negatif ataupun positif. Dengan itu maka dalam penelitian ini akan menentukan respon atau opini masyarakat terhadap Provider First Media. Menurut (Pradita Eko Prasetyo Utomo, 2019) Tanggapan negatif maupun positif dari pengguna bisa jadi dipengaruhi oleh beberapa hal yang belum menjadi perhatian dari First Media. Hal ini mungkin terjadi karena adanya beberapa faktor yang harus diperbaiki dan belum diketahui oleh pihak First Media. Dengan menggunakan text mining dapat dilihat pembicaraan apa saja yang sering dibahas oleh pengguna. Pandangan atau umpan balik masyarakat menjadi peran penting dalam peningkatan kualitas dan selalu terbukti menjadi sumber daya paling penting dan berharga perusahaan untuk mengambil langkah bisnis selanjutnya dan bahan evaluasi untuk lebih meningkatkan kualitas layanannya.

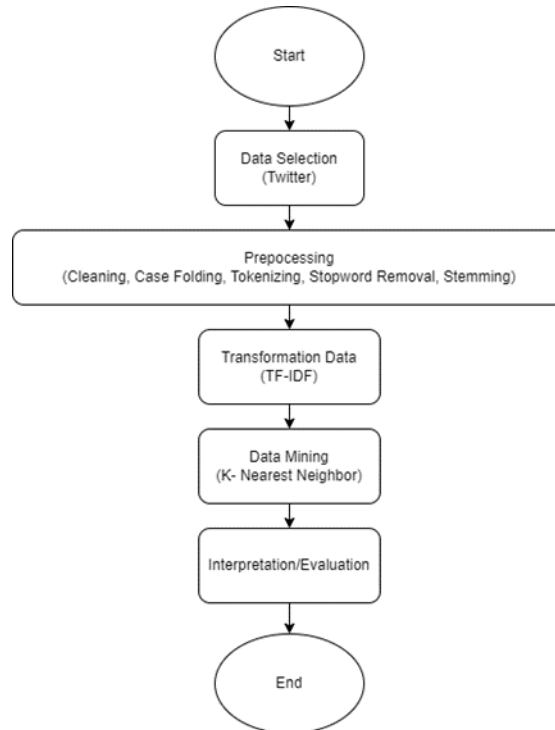
Pada penelitian kali ini akan membahas tentang sentimen pengguna media sosial twitter terhadap penyedia jasa layanan internet First Media. Dalam penelitian kali ini peneliti akan menggunakan media sosial twitter untuk mencari sebuah opini atau tanggapan dari pengguna twitter terhadap Layanan First Media. Opini atau sentimen masyarakat yang dalam istilah Twitter biasa disebut sebagai “ciutan” ini dapat berupa opini negatif maupun positif. Akan tetapi jumlah data tersebut cukup banyak sehingga dibutuhkan suatu metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentimen. Analisis sentimen merupakan sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen dikategorikan sebagai sentimen positif maupun sentimen negative. Analisis sentimen merupakan proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, dan biasanya sentimen tersebut akan diklasifikasikan menjadi opini positif dan negatif. (Saleh Hasan Badjrie, 2021).

Dengan latar belakang diatas penulis akan melakukan pengujian berupa analisis sentimen terhadap penyedia jasa layanan internet first media dengan menggunakan metode *K-Nearest Neighbor*. Penelitian ini bertujuan untuk mengetahui tanggapan terhadap kualitas dari penyedia jasa layanan internet first media dan untuk menjadi bahan evaluasi terhadap penyedia jasa layanan first media sendiri.

METODE PENELITIAN

Knowledge Discovery in Database (KDD) akan digunakan sebagai metodologi pada penelitian ini. Pada tahapan Knowledge Discovery in Database (KDD) terdiri dari 5 tahapan yaitu data selection, preprocessing data, transformation data, data mining, dan interpretation/evaluation. Pada tahapan preprocessing dilakukan case folding, tokenizing, stopword removal, dan stemming. Setelah

preprocessing dilanjutkan pada tahapan transformation data dengan pembobotan term menggunakan TF-IDF untuk meningkatkan akurasi pada hasil klasifikasi. Pada tahapan Data Mining dilakukan klasifikasi menggunakan $\text{-K-Nearest Neighbor}$. Dan pada tahapan Evaluasi akan menentukan hasil secara manual dengan menggunakan Confusion Matrix.



Gambar 1 Rancangan Penelitian

1. Data Selection

Data pada penelitian ini diperoleh dari twitter dengan menggunakan rapidminer dengan kata kunci “First Media”. Setelah mendapatkan data tweet kemudian akan dilakukan pelabelan menjadi dua kategori yaitu positif dan negatif. Pada proses pelabelan akan dibantu oleh pakar bahasa untuk meminimalisir kesalahan dalam pelabelan.

2. Preprocessing

Tahapan preprocessing ini merupakan suatu proses agar membentuk dan membersihkan data sebelum dilakukannya klasifikasi. Berikut adalah tahapan dari preprocessing:

a) Case Folding

Tahapan awal pada proses ini adalah case folding yang bertujuan untuk merubah teks menjadi huruf kecil.

b) Cleaning

Cleaning merupakan proses pembersihan data diantaranya untuk menghilangkan emoticon, ataupun simbol yang tidak diperlukan.

c) Tokenizing

Selanjutnya adalah tahapan tokenizing. Proses ini digunakan untuk melakukan pemecahan kalimat menjadi kata per kata.

d) Stemming

Proses stemming merupakan proses untuk menghilangkan imbuhan. Setelah melakukan tokenizing maka akan dilakukan penghapusan imbuhan pada hasil token.

e) Stopword Removal

Stopword removal memiliki 32 stopwords list yang akan memangkas kata pada data yang mengandung kata pada list tersebut. Stopwords list tersebut berisi kata yang tidak memiliki peran atau nilai pada proses klasifikasi.

3. Transformation Data

Pada proses transformasi adalah proses perubahan terhadap data yang telah dipilih, sehingga data tersebut menjadi sesuai dan siap digunakan untuk proses data mining. Pada tahap ini data yang telah dilakukan preprocessing akan dilakukan proses perhitungan bobot teks atau pembobotan teks menggunakan teknik TF-IDF.

4. Data Mining

Pada tahap ini akan dilakukan proses pengolahan data dan teknik klasifikasi. Data yang akan diproses merupakan data hasil yang telah ditransformasi pada tahap sebelumnya. Tahap klasifikasi adalah tahapan utama dalam proses analisis sentimen. Pada tahap ini, dilakukan implementasi metode K-Nearest Neighbor pada aplikasi RapidMiner. Dengan menggunakan metode K-Nearest Neighbor dalam penelitian ini dengan tujuan untuk mengklasifikasikan suatu objek berdasarkan atribut-atribut dan training sample. Proses klasifikasi dengan metode K-Nearest Neighbor pada aplikasi rapidminer dilakukan dengan menggunakan operator k-NN. Dengan menentukan nilai k pada metode K-Nearest Neighbor maka akan didapatkan hasil berupa sebuah akurasi.

5. Evaluasi

Pada tahap ini, setelah proses data mining dilakukan, maka data harus diberikan tampilan visual, baik berupa grafik, chart, dan lainnya agar mudah dimengerti oleh orang yang memiliki kepentingan. Pada tahap ini memerlukan perhitungan manual dengan Confusion Matriks untuk menentukan sebuah akurasinya.

HASIL DAN PEMBAHASAN

1. Data Selection

Pada tahap ini dimulai dengan pengambilan data berupa Tweet dari aplikasi Twitter dengan kata kunci "First Media". Pengambilan data ini dibantu dengan menggunakan tools Rapid Miner dengan bantuan terminal Retrive, Search Twitter, dan Select Attributes. Tweet yang diambil merupakan Tweet terbaru dan terpopuler dengan kata kunci First Media. Data yang akan diambil pada penelitian kali ini sebanyak 500 Tweet yang nantinya akan dilakukan proses selanjutnya yaitu membersihkan data – data yang tidak sesuai topik (Out Of Topic), data data yang duplicate, dan menghilangkan attribut yang tidak digunakan.

Setelah melakukan pembersihan data maka hanya tersisa 300 tweet yang nantinya akan dilakukan pelabelan. Pelabelan dilakukan dengan memberikan dua kelas yaitu positif dan negatif, pelabelan ini dibantu oleh pakar Bahasa Indonesia agar data tersebut tervalidasi. Hasil dari pelabelan data didapatkan lebih banyak sentimen negatif dari pada positif yaitu sentimen positif sebesar 90 tweet dan sentimen negatif sebesar 210 tweet

2. Preprocessing

a. Case Folding

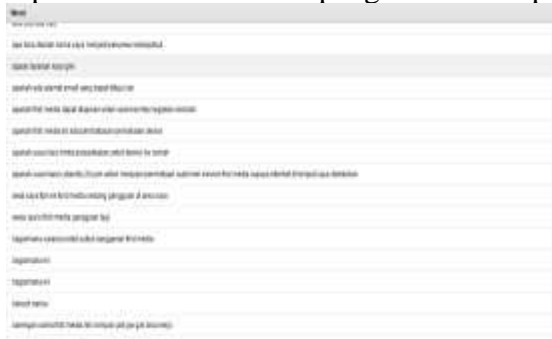
tahapan case folding merupakan tahap untuk mengubah seluruh huruf yang terdapat pada data yang telah diambil menjadi huruf kecil (lowercase). Tools Rapid Miner akan membaca data yang sebelumnya terdapat huruf kapital (uppercase) akan di ubah menjadi huruf kecil (lowercase)



Gambar 2 Hasil Case Folding

b. Cleaning

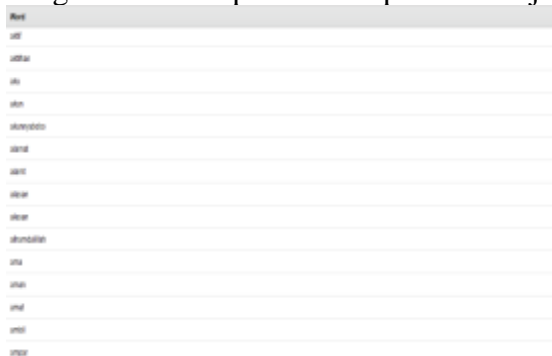
Cleaning merupakan pembersihan data diantaranya mengilangkan karakter yang tidak penting seperti emoticon, hastag, url, username ataupun symbol yang tidak diperlukan dan tidak berpengaruh dalam proses ini.



Gambar 3 Hasil Cleaning

c. Tokenizing

Tokenizing merupakan tahapan dengan memisahkan setiap kata per kata dalam sebuah kalimat atau ulasan yang ada berupa dokumen teks yang dihubungkan dengan karakter spasi atau dapat disebut juga pemecaan kalimat.



Gambar 4 Hasil Tokenizing

d. Stemming

Proses ini bertujuan untuk menghilangkan imbuhan. Imbuhan yang akan dihapus yaitu baik awalan maupun akhiran sehingga menjadi sebuah kata dasar.

Word	Atribut...	Tota... ↓	Docum...	Positif	Negatif
first	first	382	318	205	177
media	media	344	318	168	176
firstmedi...	firstmedi...	106	104	29	77
ya	ya	55	45	28	27
ga	ga	54	50	27	27
wifi	wifi	52	48	31	21
internet	internet	42	38	19	23
anjing	anjing	40	16	36	4
ngentot	ngentot	37	13	36	1
aja	aja	34	30	3	31
people	people	31	31	30	1
rumah	rumah	31	30	21	10
sinyal	sinyal	29	17	26	3

Gambar 8 Frekuensi Kemunculan Kata

Pada gambar diatas telah diketahui bahwa hasil dari TF-IDF terbentuk matriks dengan ukuran 342 X 1.229 dimana terdapat 342 dokumen dan 1.229 term. Setiap term sudah memiliki nilai TF-IDF. Nilai nol yang muncul menandakan term yang jarang muncul pada dokumen.

4. Data Mining

Pada tahapan kali ini akan dilakukan klasifikasi terhadap sentimen dari Twitter mengenai First Media dengan menggunakan algoritma K-Nearest Neighbor (KNN). Pada tahap ini dilakukan uji variasi nilai k dengan menggunakan tools Rapid Miner. Angka yang digunakan berupa angka ganjil dari 3 hingga 25.

K	Accuracy	Precission	Recall
3	84,29	95,67	72,02
5	79,14	86,66	70,29
7	79,43	82,81	75,39
9	78,29	81,37	73,76
11	76,57	78,11	74,22
13	74,57	75,14	73,63
15	76	78,80	72,45
17	74,57	78,26	69,67
19	74,86	78,87	69,64
21	73,71	76,67	69,67
23	74	77,67	69,08
25	73,14	77,96	67,42

Tabel 2 Hasil Uji Variasi Nilai K

Untuk mendapatkan hasil accuracy terbaik, maka nilai k harus semakin kecil. Hal tersebut agar mengurangi jumlah k tetangga yang kurang relevan dalam proses klasifikasi. Dapat disimpulkan bahwa nilai k dengan akurasi terbaik adalah k=3 sebesar 84%.

5. Evaluation

Hasil analisis sentimen tersebut selanjutnya dilakukan pengujian secara manual dengan menggunakan *Confusion Matrix*

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Accuracy} = \frac{169+126}{169+126+49+6} = \frac{295}{350} = 84\%$$

KESIMPULAN

Hasil pengolahan opini atau sentimen pengguna Twitter terhadap penyedia jasa layanan Internet First Media dengan menerapkan K-Nearest Neighbor yang dilakukan dengan menggunakan metodologi Knowledge Discovery in Database (KDD) menghasilkan beberapa nilai yang telah di uji variasi k nya. Nilai dengan accuracy terbaik yaitu 84% dengan nilai k=3.

DAFTAR PUSTAKA

- Abu Salam, Zeniarja, J., & Uswatun, R. (2018). ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASAEKSPEDISI BARANG J&T EKSPRESS INDONESIA). *Prosiding SINTAK*, 480.
- Aliffia Kulsumarwati, I. P. (2021). PENERAPAN SVM DAN INFORMATION GAIN PADA ANALISIS SENTIMEN PELAKSANAAN PILKADA SAAT PANDEMI. *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*.
- Badjrie, S., Pratiwi, O., & Anggana, H. (2021). ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP PRODUK INDIHOME DAN FIRST MEDIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK. *e-Proceeding of Engineering*, 9051.
- Budi, S. (2017). Text Mining Untuk Analisis Sentimen Review Film . *Techno.COM*.
- Daeli, N., & Adiwijaya. (2020). Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor. *JOURNAL OF DATA SCIENCE AND ITS APPLICATIONS*, 06.
- Dharmawan, L., Arwani, I., & Ratnawati, D. (2020). Analisis Sentimen pada Media Sosial Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 960.
- Kittipong Chomboon. (2015). An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm. *Proceedings of the 3rd International Conference on Industrial Application Engineering*, 280.
- Kusrini, & Luthfi, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Lestari, A., Perdana, R., & Fauzi, M. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1718.
- Lestari, D., & Mahdiana, D. (2021). Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis. *JURNAL INFORMATIK Edisi ke-17*, 130.

- Mi Li, & Hongpei Xu. (2018). Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technology and Health Care* 26, S513.
- Nikmatun, I., & Waspada, I. (2019). IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal SIMETRIS*, 422.
- Pratama, A. Y. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)*.
- Rezki, M., & Kholifah, D. N. (2020). Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting . *Jurnal Infortech*.
- Saifurrachman Chohan, A. N. (2020). Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naïve Bayes . *Paradigma – Jurnal Informatika dan Komputer*, P-ISSN 1410-5063, E-ISSN: 2579-3500.
- Septian, J., Fahrudin, T., & Nugroho, A. (2018). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, 43-44.
- Sikumbang, E. (2018). Penerapan Data Mining Penjualan Sepatu. *Jurnal Teknik Komputer*, 156.
- Siroj, S., Arwani, I., & Ratnawati, D. (2021). Analisis Sentimen Opini Publik pada Twitter terhadap Efek Pembelajaran Daring di Universitas Brawijaya menggunakan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3134.
- Sutoyo. (2020). Pengaruh Kualitas Layanan dan Persepsi Harga Terhadap Kepuasan Konsumen First Media DI Surabaya Timur. *JURNAL PURNA ISWARA*, 118-119.
- Suwirmayanti, N. L. (2017). Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil. *Techno.COM*, 122.
- Sya'bani, M. R. (2022). Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma. *JURIKOM (Jurnal Riset Komputer)*.
- Utomo, P., Manaar, Khaira, U., & Tri Suratno. (2019). ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA BUKALAPAK MENGGUNAKAN METODE ALGORITMA TF-IDF. (*JUSS*) *Jurnal Sains dan Sistem Informasi*, 36.
- Wahyudiyono. (2019). Implikasi Penggunaan Internet Terhadap Partisipasi Sosial. *Jurnal Komunika Jurnal Komunikasi, Media dan Informatika*, 63-64.
- Zanuardi, A., & Suprayitno, H. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan. *Jurnal Manajemen Aset Infrastruktur & Fasilitas Knowledge Discovery in Database*, 47.