

# Analisis Sentimen Mengenai Layanan Provider Indihome Berdasarkan Pendapat Pelanggan Melalui Media Sosial Twitter dengan Metode Naïve Bayes Classifier

Afif Nor Yusuf<sup>1</sup>, Endang Supriyati<sup>2</sup>, Tri Listyorini<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus

<sup>1</sup>[afifny11@gmail.com](mailto:afifny11@gmail.com)

<sup>2</sup>[endang.supriyati@umk.ac.id](mailto:endang.supriyati@umk.ac.id)

<sup>3</sup>[trilistyorini@umk.ac.id](mailto:trilistyorini@umk.ac.id)

**Abstrak**— Analisis sentimen ialah sesuatu studi yang kala ini banyak diminati oleh bermacam berbagai kalangan. Dokumen ataupun opini yang terdapat di media sosial sanggup digunakan untuk analisis sentimen. Data media sosial yang biasanya digunakan untuk analisis sentimen adalah Twitter. Sebelum suatu data diklasifikasikan, butuh dilakukan untuk proses preprocessing yang merupakan tahap untuk menghasilkan data yang lebih terstruktur. Ada pula metode yang biasanya digunakan untuk analisis sentimen adalah Naïve Bayes Classifier. Naïve Bayes Classifier bisa digunakan untuk mengategorikan suatu opini ke dalam opini positif, opini negatif dan opini netral. Tidak hanya itu, metode ini mengenakan data training dalam proses pengklasifikasian. Hasil dari metode naïve bayes sangat baik. Untuk menguji tingkatan keakuratan sistem dalam mengklasifikasi opini, sehingga pengujian memperoleh hasil klasifikasi. Hasil dari klasifikasi memperoleh hasil rata-rata 74,5%. Semakin banyak data training yang mirip dengan data testing maka hasil klasifikasi akan lebih bagus.

**Kata kunci:** Analysis Sentiment, Preprocessing, Naïve Bayes, Twitter, Text Mining

## I. PENDAHULUAN

Di masa digital seperti ini manfaat teknologi informasi sangatlah penting. Terdapat teknologi yang kala ini menjadi suatu gaya baru atau tren yaitu Big Data. Data informasi yang berbentuk teks yang memiliki kapasitas sangat besar biasa disebut dengan Big Data. Informasi berbentuk tulisan merupakan sebuah pemberitaan yang bisa didapatkan baik berasal dari bacaan, buku, ataupun media sosial itu sendiri. Tulisan merupakan suatu paparan bahasa, baik itu berbentuk ucapan maupun paparan tulisan yang mempunyai sebuah arti tertentu, berupa praktis dan berguna untuk khalayak umum dan juga keterkaitan dengan dunia ini[1].

Dengan banyaknya pemakai Twitter yang menyatakan opininya mengenai penggunaan mereka memakai provider Indihome, pandangan itu bisa digunakan buat meningkatkan

Indihome dalam menilai produk mereka, supaya bisa senantiasa bertahan dalam kompetisi dengan sebagian kompetitor lain. Pentingnya konsumen untuk sebuah industri terkait pada bagaimana industri menjalankan hubungan dengan konsumen[2].

Twitter merupakan fasilitas berjejaring sosial serta microblog daring yang membolehkan pemakainya membuat mengirim cuitan dan juga membaca catatan berbasis bacaan sampai 140 karakter, yang diketahui dengan penyebutan kicauan atau *tweet*. Jack Dorsey telah membuat Twitter pada pertengahan bulan yakni pada bulan Maret tahun 2006, pada bulan Juli sosial media Twitter dirilis. Semenjak dirilis, Twitter sudah menjadi salah satu dari 10 website yang sangat kerap didatangi di internet, serta dinamai dengan pesan pendek dari internet. Banyak pengguna Twitter yang tidak terdaftar, namun bisa membaca kicauan di Twitter. Sedangkan pengguna yang terdaftar dalam platform Twitter dapat membaca cuitan dan menuliskan cuitan melalui antar muka website[3]. Analisis sentimen mempelajari pendapat yang mengutarakan atau mengungkapkan sentimen positif, negatif atau netral. Opini mining juga disebut dengan analisis sentimen, yaitu suatu pengetahuan yang menganalisis suatu pendapat, sentimen publik terhadap entitas kelayakan produk atau jasa[4].

Pada riset tersebut dilakukannya analisis sentimen dengan memanfaatkan metode Naïve Bayes dalam pemilihan polaritas sentimen. Yang di hasil dari penelitian itu memperlihatkan ketepatan yang lumayan besar untuk metode Naïve Bayes Classifier. Berdasarkan perihal ini, pengarang berkeinginan untuk mengaplikasikan metode Naïve Naves Classifier untuk memandang sentimen publik di layanan sosial media Twitter kepada provider internet Indihome[5].

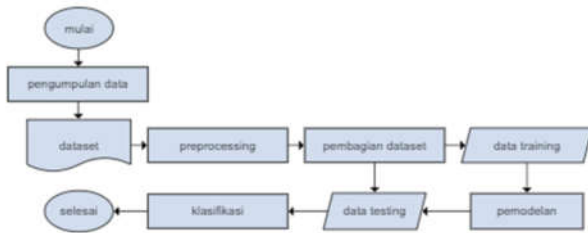
## II. METODE

Klasifikasi teks bisa didefinisikan sebagai metode untuk memastikan suatu dokumen bacaan ke dalam sesuatu golongan khusus. Pengelompokan sentimen merupakan cara memastikan

sesuatu teks ke dalam kategori positif, netral, negatif. Naïve Bayes merupakan algoritma yang kerap dipakai dalam pengkategorian bacaan. Inspirasi dasarnya merupakan mencampurkan kemungkinan kata- kata serta jenis buat memprediksi probabilitas dari jenis suatu dokumen[6].

### 2.1. Diagram Alur Proses

Riset berbentuk kumpulan fakta ataupun informasi yang dihasilkan dari sesuatu bahan riset. Dalam perihal ini materi riset yang diartikan yakni informasi *tweet* yang menyebutkan *account* Twitter Indihome yang setelah itu informasinya hendak dijadikan sebagai basis informasi. Didalam melaksanakan sebuah penelitian ini dibutuhkan beberapa alur dalam melaksanakan metode penelitian, supaya dalam metode riset berjalan dengan lancar dan sesuai dengan rencana. Gbr 1 membuktikan bagan ceruk cara riset dalam riset ini[7].



Gbr 1. Diagram Alur Proses

Pada Gbr 1. Merupakan alur proses dalam penelitian antara lain adalah tahap pengumpulan data informasi, text preprocessing, vektorisasi, pembagian data, pemodelan, serta klasifikasi.

### 2.2. Pengumpulan Data

Langkah pengumpulan data dengan melaksanakan *scraping* melalui API yang disediakan oleh Twitter dengan memanfaatkan *library tweepy* dengan bahasa pemrograman python. Data yang *scraping* merupakan berasal dari *tweet* yang tertuju pada akun Indihome dengan menggunakan kata kunci Indihome. Data yang telah *scraping* akan disimpan dalam file yang berbentuk CSV yang nantinya akan di proses lebih lanjut[7].

### 2.3. Text Preprocessing

Pada langkah ini dataset hasil dari *scraping* masih tidak struktur. Sebelum masuk proses analisa, dataset harus melalui proses *text preprocessing*[7]. Setelah data bersih maka akan dilakukan proses pelabelan secara manual. Tahap *text preprocessing* adalah sebagai berikut:

**Cleansing:** yaitu dataset akan di bersihkan dari angka, tanda baca, link url, simbol, dan emoji.

**Case Folding:** yaitu mengubah teks menjadi huruf kecil.

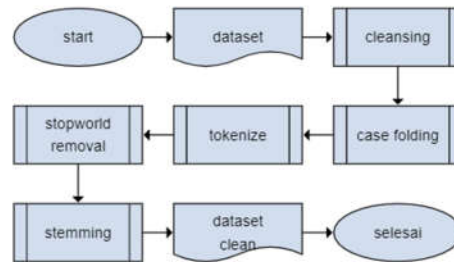
**Tokenize:** digunakan untuk mengubah kalimat menjadi kata per kata seperti "aku pulang sekolah" menjadi "aku", "pulang", "sekolah"[8].

**Stopworld:** digunakan untuk menghilangkan kata-kata yang tidak penting dalam kalimat, misalnya "adalah", "ke", "di", "dari", "dan", "dia", "kami", "aku", dan lainnya sehingga kurang relevan dengan teks[9].

TABEL 1.  
STOPWORD REMOVAL

Contoh kalimat	@IndiHome hari ini plasa telkom di pasar baru buka gak ya?
Tokenisasi	['hari', 'ini', 'plasa', 'telkom', 'di', 'pasar', 'baru', 'buka', 'gak', 'ya']
Stopworld	['plasa', 'telkom', 'pasar', 'buka']

**Stemming:** mengubah kata menjadi bentuk dasar, atau menghilangkan imbuhan pada kata. Gbr 2. berikut adalah alur *text preprocessing*.



Gbr 2. Text Preprocessing

### 2.4. Algoritma Naïve Bayes

Untuk melakukan pengelompokan teks dapat digunakannya metode Naive Bayes Classifier. Metode Naive Bayes ialah metode dalam klasifikasi probabilitas yang simpel di mana dalam metode ini dilakukan kalkulasi probabilitas dengan menggunakan metode untuk melaksanakan penjumlahan kepada frekuensi serta kombinasi nilai dari dataset[1]. Pada Persamaan (1) merupakan rumus Naive Bayes.

$$P(H|X) = \frac{P(X|H)*P(H)}{P(X)} \quad (1)$$

Keterangan:

P(H|X) : Probabilitas H berdasarkan kondisi X

P(H) : Probabilitas H.

P(X|H) : Probabilitas X berdasarkan kondisi H.

P(X) : Probabilitas X.

### III. HASIL DAN PEMBAHASAN

#### 3.1. Persiapan Data

Penelitian ini memanfaatkan data dari media sosial yang diperoleh dari Twitter, dengan metode *scraping* untuk pengambilan data *tweet*. Data yang diambil dengan kata kunci Indihome. Data yang diambil sebanyak 2.000 data *tweet* dengan metode *scraping* yang dibikin dengan pemrograman python dan *library tweepy* yang memanfaatkan API Twiter. Setelah itu data disimpan dalam format CSV untuk memudahkan mengolah data. Hasil data *scraping* bisa dilihat pada Tabel 1.

TABEL 2  
 DATA TWEET

Date	Username	Tweet	Keyword
12/9/2020 1:32:03 AM	dinakaruns	@IndiHome hari ini plasa telkom di pasar baru buka gak ya?	Indihome

#### 3.2. Text Preprocessing

Setelah pengambilan data dari Twitter selesai, langkah selanjutnya ialah *text preprocessing*. Agar dataset menjadi lebih bersih karena telah menghapus kata “RT”, username, link url, simbol, angka, emoji, dan kata-kata tidak penting. Hasil dari *text preprocessin* pada table 3.

TABEL 3  
 HASIL PREPROCESSING

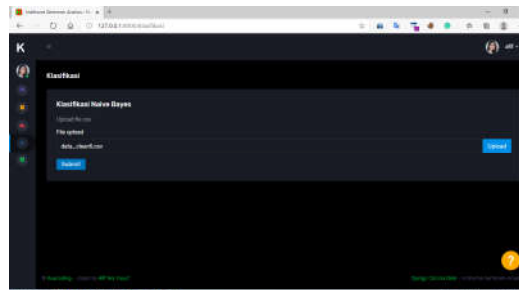
Dataset mentah	@IndiHome hari ini plasa telkom di pasar baru buka gak ya?
Hasil Text Preprocessing	plasa telkom pasar buka

#### 3.3. Proses Klasifikasi Naive Bayes

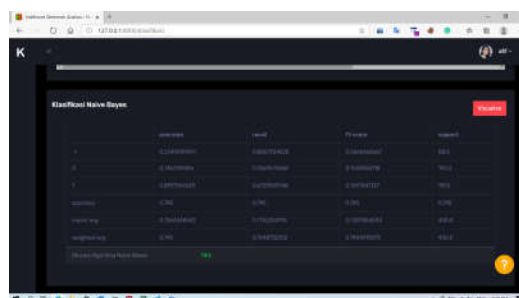
Pada tahap klasifikasi dataset yang telah diambil akan dibagi menjadi dua bagian data, yaitu 20% data *testing* dan 80% data *training*. Data *training* digunakan untuk melatih algoritma dan membuat model, sedangkan data *testing* digunakan untuk acuan dan mengetahui performa dari algoritma Naive Bayes[7]. Dengan menggunakan *library scikit-learn* untuk melakukan klasifikasi naive bayes. Data *training* yang sudah terbentuk akan dijadikan modeling untuk analisis sentimen.

Berikut ini adalah hasil dilakannya uji coba dengan antar muka yang menggunakan data sebanyak 2.000 data *tweet* yang sudah melalui proses *preprocessing* dan *labeling* terlebih dahulu. Gbr 3 adalah tampilan *user interface* metode naive

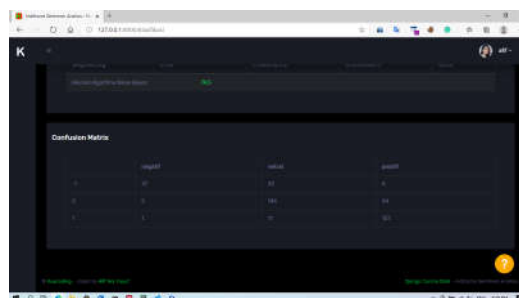
bayes untuk melakukan klasifikasi analisis sentimen dengan menggunakan *keyword* Indihome dengan cara mengunggah file CSV, dan klik button submit untuk melakukan proses klasifikasi, maka akan tampil Gbr 4. Hasil dari klasifikasi menggunakan metode naive bayes seperti pada Gbr 4.



Gbr 3. Halaman antar muka klasifikasi Naive Bayes

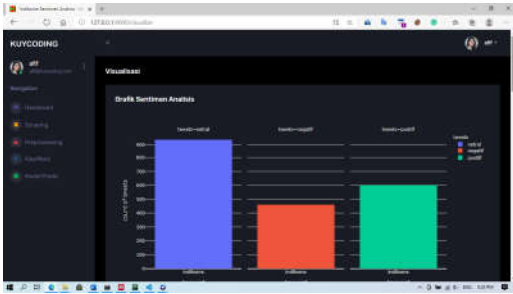


Gbr 4. Hasil Klasifikasi dengan Niave Bayes



Gbr 5. Confusion Matrix

Dari hasil klasifikasi naive bayes pada Gbr 4 didapatkan hasil akurasi 74.5 %. Gbr 5 merupakan hasil dari *confusion matrix* dari klasifikasi naive bayes. Data negatif yang sesuai dengan prediksi negatif adalah 47 data. Data netral yang masuk prediksi netral adalah 144 data. Dan data positif yang masuk prediksi positif adalah 107 data. Sedangkan data negatif yang masuk kedalam prediksi netral adalah 5 data dan 1 data prediksi positif. Sedangkan data netral yang masuk kedalam prediksi negatif adalah 33 data dan 11 data prediksi positif. Sedangkan data positif yang masuk kedalam prediksi negatif adalah 8 data dan 44 data netral. Visualisasi grafik pada Gbr 6.



Gbr 6. Visualisasi grafik Analisis Sentimen

#### IV. KESIMPULAN

Berdasarkan hasil dan pembahasan di atas, dapat disimpulkan bahwa analisis sentimen terhadap layanan provider Indihome yang didasari dengan pendapat publik di Twitter dengan menggunakan metode *Naïve Bayes Classifier* sebagai berikut:

1. Analisis sentimen terhadap data Twitter mengenai layanan Indihome yang dilakukan dengan menggunakan metode naïve bayes dengan nilai akurasi 74.5%.
2. Dari hasil penelitian yang telah dilakukan, metode naive bayes dapat diaplikasikan untuk analisis sentimen data Twitter berbahasa Indonesia.

#### REFERENSI

- [1] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [2] N. Haqqizar and T. N. Larasyanti, "Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes," no. November 2017, pp. 2017–2020, 2019.

- [3] 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.
- [4] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)," *4th Appl. Bus. Eng. Conf.*, vol. 1, no. 1, pp. 1–7, 2016.
- [5] F. Nurhuda and S. W. Sihwi, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *ITSmart J. Ilm. Teknol. dan Inf.*, vol. 2, no. 2, pp. 35–42, 2014.
- [6] N. A. Lestari and J. Timur, "METODE NAÏVE BAYES CLASSIFIER DENGAN TEXTBLOB UNTUK ANALISIS SENTIMEN TERHADAP PELAYANAN," vol. 4, no. September, 2020.
- [7] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [8] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 18, no. 2, pp. 799–806, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14744.
- [9] E. Supriyati and M. Iqbal, "Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 361–370, 2018, doi: 10.24176/simet.v9i1.1955.