



*TUGAS AKHIR*

**ANALISIS SENTIMEN PELANGGAN LAYANAN  
INTERNET TERHADAP INDIHOME TELKOM  
MENGUNAKAN ALGORITMA NAIVE BAYES  
DAN K-NEAREST NEIGHBOR**

UNIVERSITAS  
MERCU BUANA  
Divan Alif Pratama  
41518010173

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2021**



**ANALISIS SENTIMEN PELANGGAN LAYANAN  
INTERNET TERHADAP INDIHOME TELKOM  
MENGUNAKAN ALGORITMA NAIVE BAYES  
DAN K-NEAREST NEIGHBOR**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Divan Alif Pratama

41517010173

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2021

## LEMBAR PERNYATAAN ORISINALITAS

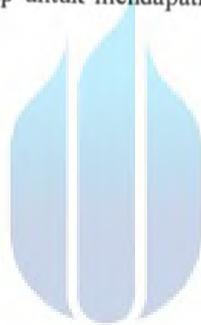
Yang bertanda tangan dibawah ini:

NIM : 41518010173

Nama : Divan Alif Pratama

Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap  
Indihome Telkom Menggunakan Algoritma Naive Bayes  
Dan K-Nearest Neighbor

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 04 Februari 2021



Divan Alif Pratama

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Divan Alif Pratama  
NIM : 41518010173  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 04 Februari 2021

UNIVERSITAS  
MERCU BUANA

A handwritten signature in blue ink is written over a rectangular stamp. The stamp contains the text 'FACULTAS ILMU KOMPUTER' and 'D7AJX00421 1937'.

Divan Alif Pratama

## SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Divan Alif Pratama  
 NIM : 41518010173  
 Judul Tugas Akhir : ANALISIS SENTIMEN PELANGGAN LAYANAN INTERNET TERHADAP INDIHOME TELKOM MENGGUNAKAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	✓
		Jurnal Nasional Terakreditasi	✓		
		Jurnal Internasional Tidak Bereputasi		Diterima	
		Jurnal Internasional Bereputasi			
Submit/dipublikasikan di :	Nama Jurnal	: Jurnal Nasional Pendidikan Teknik Informatika			
	ISSN	:2548-4265			
	Link Jurnal	:https://ejournal.undiksha.ac.id/index.php/janapati			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui  
 Dosen Pembimbing TA

Eliyani, Dr. Ir

Jakarta, 02 Agustus 2022



Divan Alif Pratama

## LEMBAR PERSETUJUAN

Nama Mahasiswa : Divan Alif Pratama  
NIM : 41518010173  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 02 Agustus 2022

Menyetujui,



Eliyani, Dr. Ir

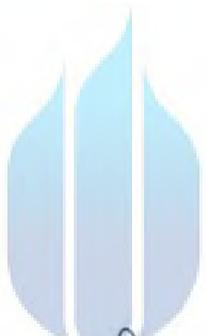
UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010173  
Nama : Divan Alif Pratama  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 2 Agustus 2022



UNIVERSITAS  
MERCU BUANA  
(Muhammad Rifqi, S.Kom, M.Kom)

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010173  
Nama : Divan Alif Pratama  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 2 Agustus 2022



## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010173  
Nama : Divan Alif Pratama  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 2 Agustus 2022



## LEMBAR PENGESAHAN

NIM : 41518010173  
Nama : Divan Alif Pratama  
Judul Tugas Akhir : Analisis Sentimen Pelanggan Layanan Internet Terhadap Indihome Telkom Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 2 Agustus 2022

Menyetujui,



Eliyani, Dr. Ir.  
Dosen Pembimbing

UNIVERSITAS

Mengetahui,

MERCU BUANA



(Wawan Gunawan, S.Kom, MT)  
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)  
Ka. Prodi Teknik Informatika

## KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Sentiment Analysis of face-to-face education in Indonesia on Twitters” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua. Ayah dan Ibu, yang tak pernah lelah mendukung dan selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik, juga tak pernah luput mendoakan yang terbaik untuk proses meraih gelar sarjana bagi saya.
2. Bapak Acmad Kodar, Drs, MT selaku Dosen Pembimbing Akademik
3. Ibu Eliyani, Dr. Ir. selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Sahabat dan kerabat, yang telah percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana dengan baik, juga tidak pernah bosan memberi dukungan dan doa.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
7. Senior atas berbagai saran yang diberikan dan junior yang dengan semangat mendukung dikala bertemu.

8. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, <tanggal-bulan-tahun>  
Divan Alif Pratama



## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR .....	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN .....	v
LEMBAR PERSETUJUAN PENGUJI .....	vi
LEMBAR PENGESAHAN .....	ix
ABSTRAK .....	x
ABSTRACT .....	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiv
NASKAH JURNAL .....	1
KERTAS KERJA.....	9
BAB 1. LITERATUR REVIEW .....	10
BAB 2. ANALISIS DAN PERANCANGAN.....	32
BAB 3. SOURCE CODE .....	34
BAB 4. DATASET.....	54
BAB 5. TAHAPAN EKSPERIMEN .....	55
BAB 6. HASIL SEMUA EKSPERIMEN.....	62
DAFTAR PUSTAKA .....	70
LAMPIRAN DOKUMEN HAKI.....	73
LAMPIRAN KORESPONDENSI .....	75



**ANALISIS SENTIMEN PELANGGAN LAYANAN INTERNET  
TERHADAP INDIHOME TELKOM MENGGUNAKAN ALGORITMA  
NAIVE BAYES DAN K-NEAREST NEIGHBOR**

Divan Alif Pratama<sup>1</sup>, Eliyani<sup>2</sup>

Teknik Informatika, Ilmu Komputer, Universitas Mercu Buana  
<sup>1</sup>41518010173@student.mercubuana.ac.id, <sup>2</sup>eliyani@mercubuana.ac.id

**Abstrak**

Seiring berjalannya waktu, kemajuan teknologi informasi semakin berkembang pesat dan merupakan satu sarana komunikasi yang membuat penggunaannya dapat dengan mudah berpartisipasi dan berbagi. Bidang ini melakukan studi kasus mengenai opini masyarakat, sentimen, evaluasi, layanan, organisasi, individu, permasalahan topik, dan lainnya. Permasalahan yang ada didalam pelayanan ISP masih banyak yang masih dirasa kurang memuaskan dari para pelanggan. Dari mulai kecepatan internet, gangguan teknis, pelayanan, dan lainnya. Internet Service Provider (ISP) secara umum adalah produsen atau lembaga yang memberikan pelayanan kepada konsumen supaya bisa mengakses internet. Dari permasalahan tersebut, penulis mencoba meneliti mengenai analisa atau pendapat masyarakat terhadap pelayanan provider Indihome. Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma K-Nearest Neighbor (KNN) memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage split pada dua model data yang diuji yaitu Indihome dan Firstmedia dengan menggunakan labeling otomatis dan manual, bahwa eksperimen dengan menggunakan labeling manual lebih representatif dalam menentukan kelas data. Model percentage split 80:20 memiliki nilai akurasi terbaik pada data yang diuji baik Indihome maupun Indihome dengan nilai akurasi 87% dan 94% pada eksperimen labeling otomatis dan 83% dan 63% pada eksperimen labeling manual. Proporsi data 80:20 dari dataset pelatihan dan pengujian memiliki hasil yang lebih baik.

**Kata kunci:** Naive Bayes, K-Nearest Neighbor (KNN), Twitter, indihome, Rapidminer

**Abstract**

Over time, advances in information technology are growing rapidly and are a means of communication that allows users to easily participate and share. This field conducts case studies on public opinion, sentiment, evaluation, services, organizations, individuals, topic issues, and others. The problems that exist in ISP services are still many that are still considered unsatisfactory from customers. Starting from internet speed, technical problems, services, and others. Internet Service Providers (ISPs) in general are producers or institutions that provide services to consumers in order to access the internet. From these problems, the author tries to examine the analysis or public opinion on the Indihome service provider. The results of this study are experiments using the K-Nearest Neighbor (KNN) algorithm have the highest accuracy values in the three percentage split experimental scenarios on the two data models tested, namely Indihome and Firstmedia using automatic and manual labeling, that experiments using manual labeling more representative in determining the data class. The 80:20 percentage split model has the best accuracy value on the data tested for both Indihome and Indihome with an accuracy value of 87% and 94% in the automatic labeling experiment and 83% and 63% in the manual labeling experiment. The 80:20 data proportion of the training and testing datasets had better results.

**Keywords :** Naive Bayes, K-Nearest Neighbor (KNN), Twitter, indihome, Rapidminer

Diterima Redaksi: dd-mm-2021 | Selesai Revisi: dd-mm-2021 | Diterbitkan Online: dd-mm-2021  
DOI: <https://doi.org/10.23887/janapati.v10i1.xxx>

**PENDAHULUAN**

Seiring berjalannya waktu, kemajuan teknologi informasi semakin berkembang pesat dan merupakan satu sarana komunikasi yang membuat penggunaannya dapat dengan mudah berpartisipasi dan

berbagi. Teknologi informasi didalam kehidupan sehari - hari, sangat berguna sebagai tempat bertukar pikiran jarak jauh, seperti di internet ataupun website. Dalam hal ini, masyarakat mengkritik mengenai

pelayanan dari provider layanan internet Indihome yang ada di Indonesia [1].

Analisa sentimen adalah salah satu cabang ilmu dari textmining, natural language program dan artificial intelligence. Bidang ini melakukan studi kasus mengenai opini masyarakat, sentimen, evaluasi, layanan, organisasi, individu, permasalahan topik, dan lainnya.

Salah satu media sosial dengan jumlah pengguna aktif mencapai 15,7 juta orang pada di negara Indonesia adalah Twitter. Twitter adalah media umum yang digunakan buat mencari informasi perihal usaha, hiburan, ekonomi, politik, dan lainnya. dengan banyaknya jumlah tweet yang dipublikasikan melalui Twitter, tweet tadi dapat mengandung opini pengguna mengenai sebuah objek, objek tersebut bisa berupa peristiwa disekitar rakyat mirip suatu produk atau layanan. Hal ini membuat perusahaan indihome menyebar luaskan informasi seputar contoh harga provider dan layanan indihome [2].

Dengan dijamin sekarang membutuhkan koneksi internet yang sangat kencang di PT. Telekomunikasi Indonesia (PT.TELKOM) merupakan salah satu perusahaan milik negara yang bergerak di bidang penyedia jasa layanan telekomunikasi yang mana menerapkan teknologi telekomunikasi digital pada perangkat system telekomunikasi yang dimilikinya[3].

Penelitian lain mengenai analisis sentimen dengan opini pengguna layanan Telkom yang dilakukan oleh Haranto tahun 2019. Penelitian tersebut bertujuan untuk mengetahui persentase sentimen positif dan negatif dari Telkom naïve bayes, memiliki hasil pengujian yaitu accuracy 79,6%, precision 76,5%, recall 72,8% , dan F1-score 74,6% untuk Telkom. Analisis sentimen ialah suatu metode dalam memahami, mengekstrasi data sentimen yang biasanya akan dikategorikan berdasarkan polaritasnya[4].

Penelitian analisis sentimen dengan objek opini maskapai penerbangan pada Twitter yang dilakukan oleh Pravina tahun 2019[5]. Penelitian tersebut bertujuan untuk mengklasifikasikan komentar masyarakat ke dalam kelas positif atau kelas negatif, agar dapat membantu pihak perusahaan maupun

masyarakat, dengan metode Support Vector Machine penelitian tersebut memiliki hasil accuracy 40%, precision 40%, recall 100%, dan 57,14% f-measure[2].

Permasalahan yang ada didalam pelayanan ISP masih banyak yang masih dirasa kurang memuaskan dari para pelanggan. Dari mulai kecepatan internet, gangguan teknis, pelayanan, dan lainnya. Internet Service Provider (ISP) secara umum adalah produsen atau lembaga yang memberikan pelayanan kepada konsumen supaya bisa mengakses internet[6]. Dari permasalahan tersebut, penulis mencoba meneliti mengenai analisa atau pendapat masyarakat terhadap pelayanan provider Indihome.

## METODE

### 2.3.1 Jenis Penelitian.

Jenis penelitian ini termasuk dalam sifat kuantitatif atau dengan metode penelitian kuantitatif. Jenis penelitian ini berbentuk angka atau menggambarkan fenomena atau fakta penelitian secara apa adanya atau menjelaskan fenomena yang terjadi di masyarakat secara mendalam dengan cara mengumpulkan data secara dalam dan lengkap. Dalam penelitian ini kelengkapan dan kedalaman data yang diteliti merupakan suatu yang sangat penting[6]. Penelitian ini pengumpulan datanya melalui pengambilan komentar-komentar yang diberikan masyarakat terhadap suatu tweet di postingan twitter yang dapat disimpan dalam bentuk format .xlsx atau .csv[7]. Penelitian ini menggunakan 2.000 data tweet.

### 2.3.2 Metode Pengumpulan Data.

Metode pengumpulan data yang digunakan untuk penelitian merupakan pengambilan data-data tweet dari Twitter dapat dilakukan dengan mengintegrasikan API Twitter, pengambilan komentar pelanggan secara manual yang dilakukan berdasarkan query atas term objek pada aplikasi pengolah data mining yang terhubung dengan API Twitter[8]. Adapun data tweet yang akan diperoleh berisi text atau komentar mengenai opini atau tanggapan masyarakat mengenai pelayanan yang

diberikan dari beberapa Internet Service Provider yang ada di Indonesia, diantaranya ialah Indihome[9].

### 2.3.3 Tahap Penelitian.

Proses sentimen analisis pada kepuasan pelanggan terhadap pelayanan Internet Service Provider yang ada di Indonesia berdasarkan komentar publik di Twitter dapat dilihat pada Gambar 1.

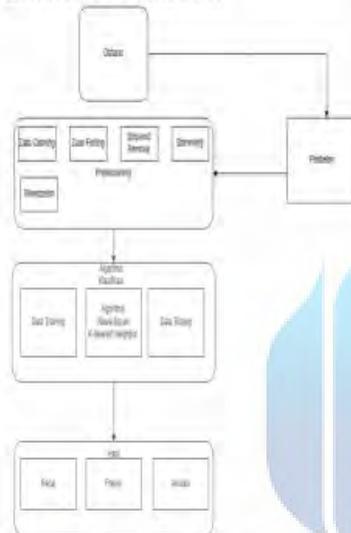


Diagram 1 Tahapan Penelitian Tahap yang akan dilakukan pada Penelitian, sebagai berikut :

#### 2.3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan cara pengambilan data yang memanfaatkan API Twitter, dilakukan dengan mengintegrasikan API Twitter, komentar pelanggan pada Twitter Indihome dan Firstmedia yang dilakukan dengan cara scraping berdasarkan query atas term objek pada aplikasi pengolahan data mining yang terhubung dengan API Twitter[10]. Proses pengumpulan data dimulai dari mengatur Twitter API di Rapid Miner Studio. Selanjutnya Twitter API dihubungkan ke internet untuk login Twitter. Langkah berikutnya adalah memasukkan query dan memilih bahasa Indonesia setelah itu dijalankan dan data akan tampil sesuai hashtag tersebut. Data dalam Rapid Miner Studio disimpan ke Microsoft Excel untuk dilakukan proses selanjutnya[11]. Tweet dan komentar yang akan diperoleh berisi mengenai tingkat kepuasan masyarakat mengenai pelayanan Internet Service Provider yang ada di Indonesia[12].

#### 2.3.2 Pengolahan Data Awal.

Sebelum diproses, perlu dilakukan pengolahan data yang diperoleh dari twitter. Proses pengolahan data atau text preprocessing berfungsi untuk merubah data teks yang tidak terstruktur menjadi data yang terstruktur.

Berikut ini adalah penjelasan dari tahapan-tahapan preprocessing tersebut:

1. Data Cleansing adalah tahap di mana karakter dan tanda baca yang tidak diperlukan dihapus dari teks. Cleansing bertujuan untuk mengurangi gangguan/noise pada dataset. Tahap-tahap pada cleansing ini adalah Hapus URL, Hapus Username, Hapus tanda RT dan Hapus Hashtag[13].

2. Case Folding = Transform case merupakan operator untuk case folding. Case folding yaitu merubah bentuk kata menjadi bentuk dasarnya agar sebuah karakter dapat seragam (lower case)[10]. Case folding dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. oleh karena itu peran case folding dibutuhkan dalam mengonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil[14].

3. Tokenization merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna. Tokenisasi memisah setiap kata yang menyusun kalimat[15].

4. Stopword Removal merupakan proses menghilangkan kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan. Kata yang dihilangkan adalah kata yang ada dalam dictionary yang berisi daftar kata (stopword list)[16].

5. Stemming merupakan tahapan untuk melakukan proses mengubah kata yang berimbuhan infix maupun suffix menjadi sebuah kata dasar yang akan lebih mengandung sebuah makna untuk peroleh suatu informasi sehingga komentar akan menjadi lebih spesifik dalam pengkategorian[17].

6. Normalisasi text adalah tahap normalisasi teks. Pada tahap ini semua teks kalimat diubah menjadi teks yang secara lengkap memperlihatkan cara pengucapannya. Normalisasi teks meliputi perubahan singkatan, akronim, angka, tanggal, waktu, karakter-karakter khusus, dan simbol-simbol dengan bentuk huruf alphabet lengkap sehingga tidak terjadi ambiguitas berkenaan dengan cara pengucapan.

#### 2.3.3 Preprocessing

Setelah melakukan tahap preprocessing, data yang telah dipreprocessing menjadi 2057 untuk data provider Indihome dan 1949 untuk data provider Firstmedia. Berikut merupakan hasil



Tabel 2 Naive Bayes Indihome Split 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	89%	54%	87%
1	87%	98%	87%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 80% : 20% bahwa untuk studi kasus Indihome memiliki nilai akurasi 86% untuk algoritma KNN dan 94% untuk algoritma Naive Bayes. Berikut merupakan hasil dari kinerja persentasi split 80% : 20%.

Tabel 3 KNN Indihome Split 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	88%	30%	82%
1	81%	99%	82%

Tabel 4 Naive Bayes Indihome Split 80:20 Labeling Otomatis.

	Precision	Recall	Accuracy
0	87%	51%	95%
1	85%	97%	95%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 70% : 30% bahwa untuk studi kasus Indihome memiliki nilai akurasi 98% untuk algoritma KNN dan 83% untuk algoritma Naive Bayes.

Tabel 5 KNN Indihome Split 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	94%	67%	91%
1	90%	99%	91%

Tabel 6 Naive Bayes Indihome Split 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	86%	46%	83%
1	83%	97%	83%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 90% : 10% bahwa untuk studi kasus Indihome memiliki nilai akurasi 72% untuk algoritma KNN dan 67% untuk algoritma Naive Bayes.

Tabel 7 KNN Indihome 90:10 Labeling Manual

	Precision	Recall	Accuracy
0	95%	78%	86%
1	80%	95%	86%

Tabel 8 Naive Bayes Indihome 90:10 Labeling Manual

	Precision	Recall	Accuracy
0	64%	70%	65%
1	66%	59%	65%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 80% : 20% bahwa untuk studi kasus Indihome memiliki nilai akurasi 82% untuk algoritma KNN dan 63% untuk algoritma Naive Bayes.

Tabel 9 KNN Indihome 80:20 Labeling Manual

	Precision	Recall	Accuracy
0	69%	99%	76%
1	97%	53%	76%

Tabel 10 Naive Bayes Indihome 80:20 Labeling Manual.

	Precision	Recall	Accuracy
0	67%	74%	67%
1	67%	60%	67%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 70% : 30% bahwa untuk studi kasus Indihome memiliki nilai akurasi 97% untuk algoritma KNN dan 57% untuk algoritma Naive Bayes.

Tabel 11 KNN Indihome 70:30 Labeling Manual

	Precision	Recall	Accuracy
0	100%	84%	92%
1	85%	100%	92%

Tabel 12 Naive Bayes Indihome 70:30 Labeling Manual

	Precision	Recall	Accuracy
0	67%	70%	65%
1	62%	58%	65%

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma K-Nearest

Neighbor (KNN) memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage split pada dua model data yang diuji yaitu Indihome dan Firstmedia dengan menggunakan labeling otomatis dan manual. Hasil dari pembahasan terangkum dalam tabel berikut:  
Percobaan Indihome.

Tabel 13 Perbandingan Akurasi Labeling Otomatis

Percobaan	Indihome	
	KNN	NB
90:10	95%	87%
80:20	82%	95%
70:30	91%	83%

Tabel 14 Perbandingan Akurasi Labeling Manual

Percobaan	Indihome	
	KNN	NB
90:10	86%	65%
80:20	76%	67%
70:30	92%	65%

Dari tabel berikut disimpulkan bahwa eksperimen dengan menggunakan labeling manual lebih representatif dalam menentukan kelas data. Model percentage split 80:20 memiliki nilai akurasi terbaik pada data yang diuji baik Indihome maupun Indihome dengan nilai akurasi 82% dan 85% pada eksperimen labeling otomatis dan 76% dan 67% pada eksperimen labeling manual. Proporsi data 80:20 dari dataset pelatihan dan pengujian memiliki hasil yang lebih baik karena memberikan nilai evaluasi yang mendekati keseimbangan setelah proses penyetulan. Percobaan ini menunjukkan bahwa semakin banyak jumlah training dataset maka akan semakin baik nilai evaluasi yang akan diperoleh karena akan banyak proses pembelajaran yang terjadi pada training dataset. Sehingga bisa disimpulkan model percentage split 80:20 dengan menggunakan algoritma K-Nearest Neighbor merupakan skenario eksperimen yang paling baik pada penelitian ini.

#### Kesimpulan

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma K-Nearest Neighbor (KNN) memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage split pada dua model data yang diuji yaitu Indihome dan Firstmedia dengan menggunakan labeling otomatis dan manual, bahwa eksperimen dengan menggunakan labeling manual lebih representatif dalam menentukan kelas data. Model percentage split 80:20 memiliki nilai akurasi terbaik pada data yang diuji baik Indihome maupun Indihome dengan nilai akurasi 87% dan 94% pada eksperimen labeling otomatis dan 83% dan 63% pada eksperimen labeling manual. Proporsi data 80:20 dari dataset pelatihan dan pengujian memiliki hasil yang lebih baik.

#### Daftar Rujukan

- [1] R. Juwita, "Media Sosial dan Perkembangan Komunikasi Korporat," *J. Penelit. Komun.*, vol. 20, no. 1, pp. 47–60, 2017, doi: 10.20422/jpk.v20i1.136.
- [2] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Pros. SNST*, vol. 7, pp. 37–42, 2016, [Online]. Available: [https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING\\_SNST\\_FT/article/view/1506/1589](https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/1506/1589).
- [3] S. Sabloak, J. Wijaya, A. Rahman, and M. Arman, "ANALISIS PEMANTAUAN LAN MENGGUNAKAN METODE QoS DAN PENGKLASIFIKASIAN STATUS JARINGAN INTERNET MENGGUNAKAN ALGORITMA NAIVE BAYES," *J. Ilm. Teknol. Infomasi Terap.*, vol. 4, no. 2, pp. 131–140, 2018, doi: 10.33197/jitter.vol4.iss2.2018.159.
- [4] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine ( SVM )," vol. 4, pp. 650–658, 2020,

- doi: 10.30865/mib.v4i3.2181.
- [5] M. W. Pertiwi, "Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, K-NN dan SVM," *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 27–32, 2019.
- [6] M. Kusmira, "Analisis Sentimen Registrasi Ulang Kartu SIM pada Twitter Menggunakan Algoritma SVM dan K-NN," *INTI Nusa Mandiri*, vol. 14, no. 1, pp. 105–110, 2019.
- [7] G. Li, Q. S. Zheng, L. Zhang, S. Z. Guo, and L. Y. Niu, "Sentiment Information based Model for Chinese text Sentiment Analysis," *2020 IEEE 3rd Int. Conf. Autom. Electron. Electr. Eng. AUTEEE 2020*, pp. 366–371, 2020, doi: 10.1109/AUTEEE50969.2020.9315668.
- [8] E. M. O. N. Taofik Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, pp. 32–37, 2021, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>.
- [9] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.
- [10] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [11] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [12] A. Rohanah, D. L. Rianti, and B. N. Sari, "Perbandingan Naïve Bayes dan Support Vector Machine untuk Klasifikasi Ulasan Pelanggan Indihome," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 23, 2021, doi: 10.30998/string.v6i1.9232.
- [13] T. Meisya *et al.*, "PERBANDINGAN KERNEL SUPPORT VECTOR MACHINE ( SVM ) DALAM PENERAPAN ANALISIS SENTIMEN VAKSINISASI COVID-19," vol. 4, no. 2, pp. 139–145, 2021.
- [14] A. Poomima and K. S. Priya, "A Comparative Sentiment Analysis of Sentence Embedding Using Machine Learning Techniques," *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 493–496, 2020, doi: 10.1109/ICACCS48705.2020.9074312.
- [15] S. Birogul, G. Temur, and U. Kose, "YOLO Object Recognition Algorithm and 'Buy-Sell Decision' Model over 2D Candlestick Charts," *IEEE Access*, vol. 8, pp. 91894–91915, 2020, doi: 10.1109/ACCESS.2020.2994282.
- [16] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," 2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019, pp. 593–596, 2019, doi: 10.1109/ICACTM.2019.8776800.
- [17] A. Al Bataineh, "A comparative analysis of nonlinear machine learning algorithms for breast cancer detection," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 3, pp. 248–254, 2019, doi: 10.18178/ijmlc.2019.9.3.794.



- [18] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
- [19] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 4763, 2018, doi: 10.11591/ijece.v8i6.pp4763-4771.
- [20] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.

UNIVERSITAS  
MERCU BUANA

## **KERTAS KERJA**

### **Ringkasan**

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Analisa Sentimen Pelanggan Layanan Internet Terhadap Komentar Di Media Sosial Twitter Indihome Telkom Menggunakan Metode Naive Bayes dan K-Nearest Neighbor (KNN)”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS  
MERCU BUANA