

**K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK  
MENGANALISIS KEPUASAN PELANGGAN PROVIDER  
INTERNET FIRST MEDIA PADA TWITTER**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Teknik Informatika



disusun oleh

**AAS ANDRI KURNIAWAN**

**17.11.1532**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK  
MENGANALISIS KEPUASAN PELANGGAN PROVIDER  
INTERNET FIRST MEDIA PADA TWITTER**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**AAS ANDRI KURNIAWAN**

**17.11.1532**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**SKRIPSI**

**K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK  
MENGANALISIS KEPUASAN PELANGGAN PROVIDER  
INTERNET FIRST MEDIA PADA TWITTER**

yang disusun dan diajukan oleh

**AAS ANDRI KURNIAWAN**

**17.11.1532**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal <24 Juni 2023>

**Dosen Pembimbing,**



**Yoga Prisyanto, S.kom., M.Eng**

**NIK. 190302412**

HALAMAN PENGESAHAN

SKRIPSI

**K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK  
MENGANALISIS KEPUASAN PELANGGAN PROVIDER  
INTERNET FIRST MEDIA PADA TWITTER**

yang disusun dan diajukan oleh

**Aas Andri Kurniawan**

**17.11.1532**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal <24 Juni 2023>

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Erni Seniwati, S.Kom, M.Cs**  
NIK. 190302231



---

**Dina Maulina, M.Kom**  
NIK. 190302250



---

**Harvoko, S.Kom, M.Cs**  
NIK. 190302286



---

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal < 24 Juni 2023 >

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom.**  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : AAS ANDRI KURNIAWAN  
NIM : 17.11.1532

Menyatakan bahwa Skripsi dengan judul berikut:

### **K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK MENGANALISIS KEPUASAN PELANGGAN PROVIDER INTERNET FIRST MEDIA PADA TWITTER**

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 24 Juni 2023

Yang Menyatakan,



AAS ANDRI KURNIAWAN

## HALAMAN PERSEMBAHAN

Allhamdulillah penulis Panjatkan pujisyukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya, sehingga diberi kesempatan untuk dapat menyelesaikan skripsi ini dengan sebaik-baiknya dengan segala kekurangan penulis. Segala syukur penulis ucapkan kepada-Mu karena telah menghadirkan mereka yang memberikan semangat dan doa disaat menjalani proses pembuatan skripsi ini. Dengan segala kerendahan hati saya persembahkan skripsi ini kepada :

1. Kedua Orang Tua yang selalu mendoakan, memberi semangat serta motivasi supaya dapat menyelesaikan skripsi ini dengan lancar serta bermanfaat bagi semua.
2. Bapak. Yoga Prisyanto, S.Kom, M.Eng. selaku dosen pembimbing dalam skripsi ini yang tidak lelah untuk tetap membimbing dan mengingatkan penulis dari awal hingga akhir proses pembuatan skripsi.
3. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu baik ilmu akademik maupun ilmu non-akademik selama kuliah.
4. Keluarga besar kelas 17-SIIF-09 yang telah bersama-sama menemani selama kuliah. Semoga silaturahmi kita tetap terjaga.
5. Serta orang-orang yang selalu membantu peneliti dalam mengerjakan skripsi yang tidak bisa disebut namanya satu-persatu.

Saya ucapkan terima kasih yang sebesar-besarnya untuk kalian semua. Sukses untuk kalian semua, semoga Allah SWT memberikan rahmat dan hidayah-Nya kepada kita semua. Dan semoga skripsi ini dapat bermanfaat dan berguna untuk kemajuan ilmu pengetahuan kedepannya.

## KATA PENGANTAR

Puji dan syukur saya panjatkan kehadiran Allah S.W.T

Alhamdulillah penulis panjatkan puji syukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi yang berjudul “K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK MENGANALISIS KEPUASAN PELANGGAN PROVIDER INTERNET FIRST MEDIA PADA TWITTER”.

Selama proses pengerjaan skripsi ini penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak dan berkah dari Allah SWT sehingga kendala-kendala tersebut bisa diatasi. Selanjutnya ucapan trima kasih penulis sampaikan kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak. Yoga Prisyanto, S.Kom, M.Eng. selaku dosen pembimbing yang telah memberikan banyak masukan yang membantu membimbing dalam menyelesaikan skripsi ini
3. Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Windha Mega D, M.Kom selaku Kepala Prodi Informatika Universitas Amikom Yogyakarta.
5. Dosen Penguji (Ibu Erni Seniwati, S.Kom, M.Cs, Ibu Dina Maulina, M.Kom, Bpk Haryoko, S.Kom, M.Cs) yang telah memberikan masukan terhadap penelitian ini.
6. Kedua orang tua dan keluarga yang selalu memberikan doa, dukungandan semangat.

Yogyakarta, <24 Juni 2023>

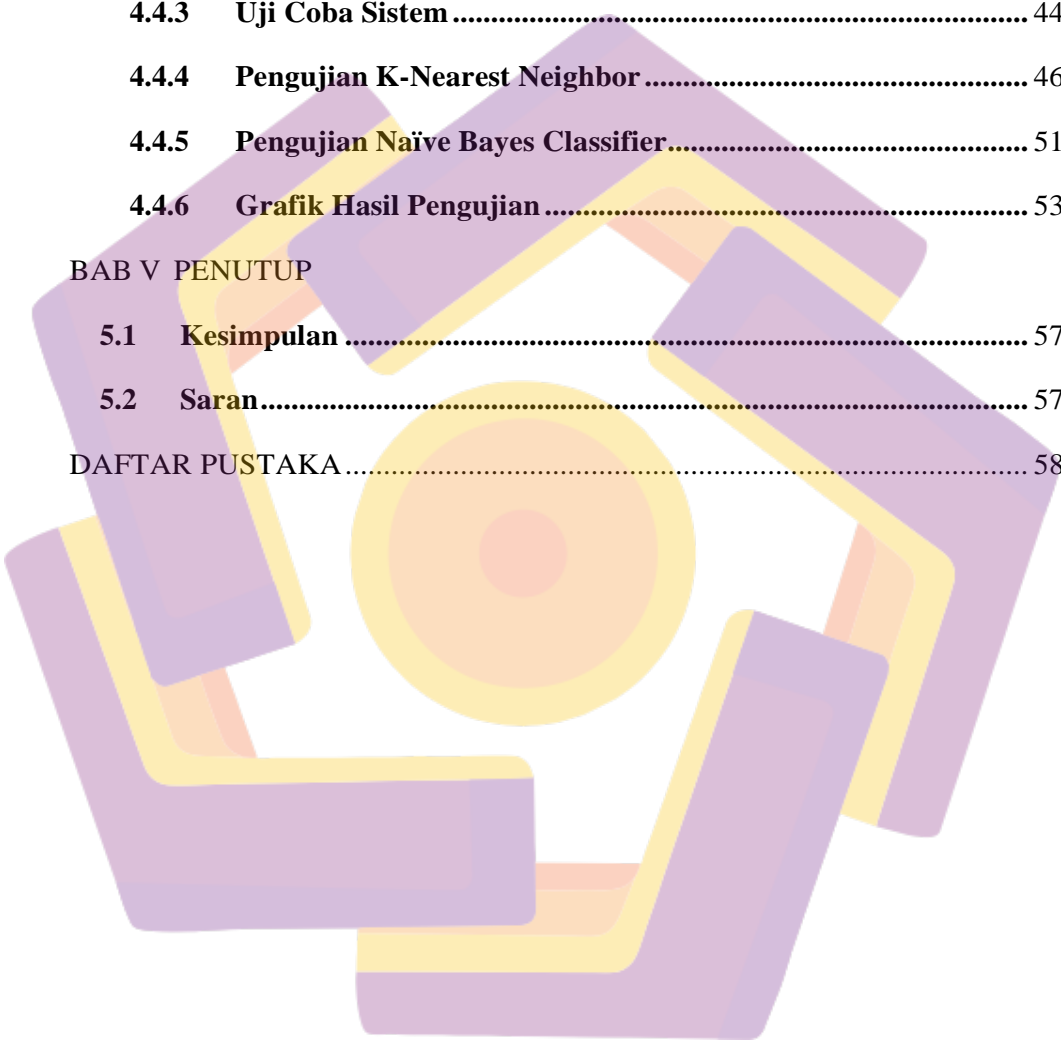
Penulis

## DAFTAR ISI

DAFTAR ISI .....	i
DAFTAR TABEL .....	iv
DAFTAR GAMBAR .....	vi
INTISARI .....	vii
ABSTRACT .....	viii
BAB I PENDAHULUAN	
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	3
<b>1.3 Batasan Masalah</b> .....	3
<b>1.4 Tujuan Penelitian</b> .....	4
<b>1.5 Manfaat Penelitian</b> .....	4
<b>1.6 Metode Penelitian</b> .....	4
<b>1.7 Sistematika Penulisan</b> .....	5
BAB II LANDASAN TEORI	
<b>2.1 Analisis Sentimen</b> .....	7
<b>2.2 Tinjauan Pustaka</b> .....	8
<b>2.2.1 Penelitian Terdahulu</b> .....	8
<b>2.3 Landasan Teori</b> .....	11
<b>2.3.1 Twitter API</b> .....	11
<b>2.3.2 Text Mining</b> .....	12
<b>2.3.3 Preprocessing</b> .....	13
<b>2.3.4 Cleansing</b> .....	13



2.3.5	<b>Case Folding</b> .....	13
2.3.6	<b>Tokenizing</b> .....	14
2.3.7	<b>Stopword Removal</b> .....	14
2.3.8	<b>Stemming</b> .....	15
2.3.9	<b>Algoritma Nazief dan Adriani</b> .....	16
2.3.10	<b>Pembobotan Kata TF-IDF</b> .....	17
2.3.11	<b>Klasifikasi</b> .....	18
2.3.12	<b>K-Nearest Neighbor</b> .....	18
2.3.13	<b>Naïve Bayes</b> .....	19
<b>BAB III METODE PENELITIAN</b>		
3.1	<b>Deskripsi Umum</b> .....	21
3.2	<b>Analisis Kebutuhan Sistem</b> .....	22
3.3	<b>Tahap Pengumpulan Data</b> .....	23
3.3.1	<b>Daftar Pustaka</b> .....	23
3.3.2	<b>Crawling Data Twitter</b> .....	23
3.4	<b>Tahap Preprocessing</b> .....	24
3.5	<b>Pembobotan TF-IDF</b> .....	26
3.6	<b>Tahap Implementasi dan Evaluasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor</b> .....	27
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	<b>Proses Pengumpulan Data</b> .....	37
4.2	<b>Preprocessing Data</b> .....	37
4.2.1	<b>Proses Case Folding</b> .....	38
4.2.2	<b>Proses Tokenizing</b> .....	38
4.2.3	<b>Proses Filtering</b> .....	39

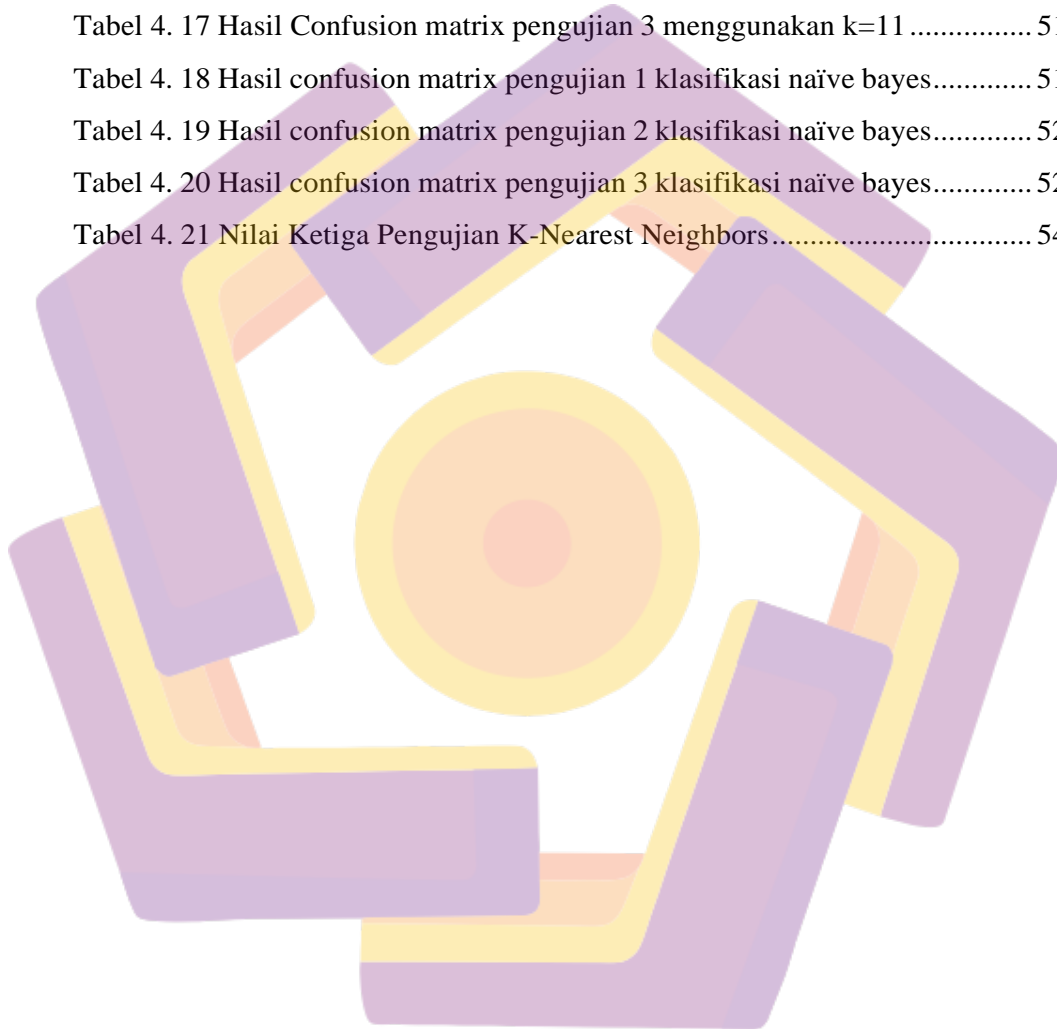


4.2.4	Stemming .....	40
4.3	Pembobotan TF-IDF .....	41
4.4	Implementasi Klasifikasi .....	41
4.4.1	Klasifikasi data menggunakan Textblob.....	41
4.4.2	Pembobotan TF-IDF .....	44
4.4.3	Uji Coba Sistem .....	44
4.4.4	Pengujian K-Nearest Neighbor .....	46
4.4.5	Pengujian Naïve Bayes Classifier.....	51
4.4.6	Grafik Hasil Pengujian .....	53
BAB V PENUTUP		
5.1	Kesimpulan .....	57
5.2	Saran.....	57
DAFTAR PUSTAKA.....		58

## DAFTAR TABEL

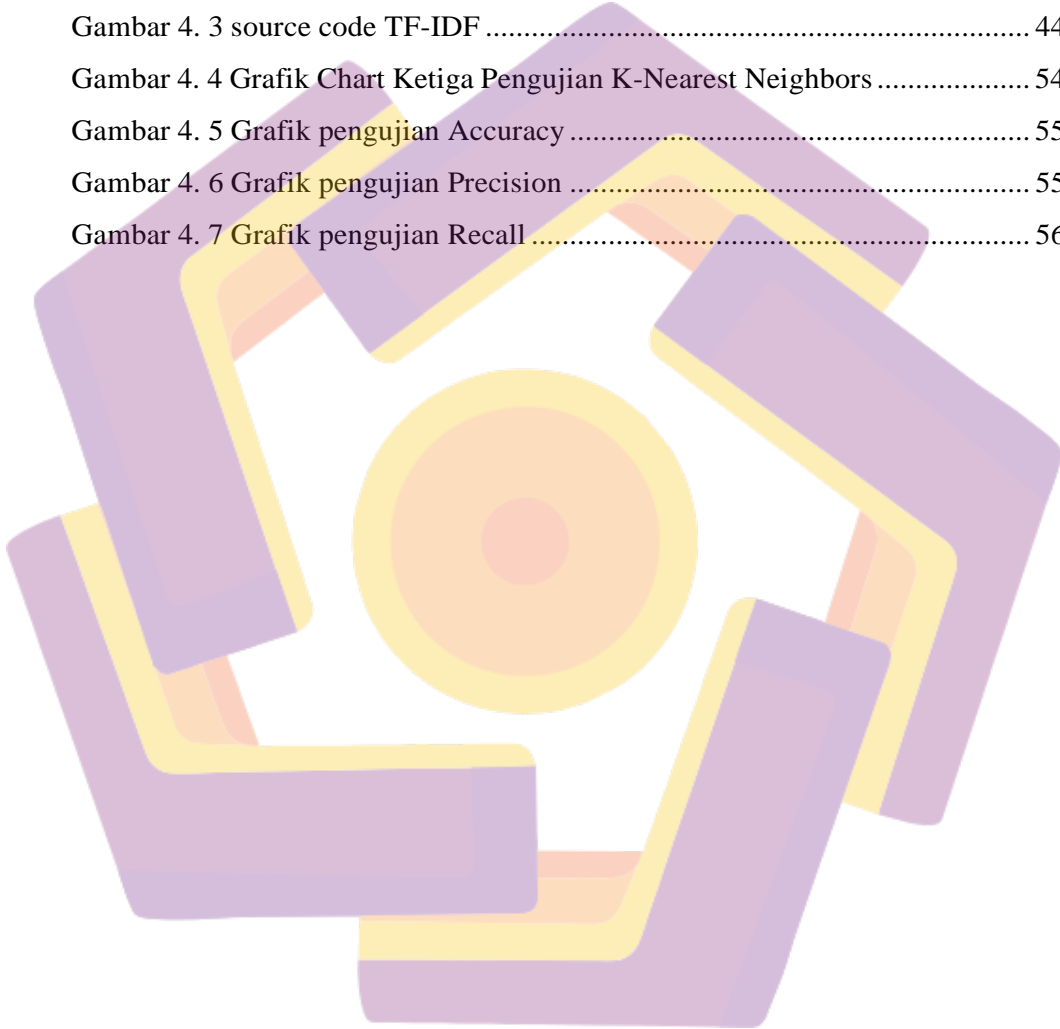
Tabel 2. 1 Studi Literatur Sejenis .....	10
Tabel 2. 2 Contoh Case Folding .....	13
Tabel 2. 3 Contoh Tokenizing .....	14
Tabel 2. 4 Contoh Stopword Removal.....	14
Tabel 2. 5 Contoh Stemming .....	16
Tabel 3. 1 Proses Case Folding.....	24
Tabel 3. 2 Proses Tokenizing.....	25
Tabel 3. 3 Proses Filtering .....	25
Tabel 3. 4 Proses Stemming .....	26
Tabel 3. 5 Proses TF-IDF .....	27
Tabel 3. 6 Data Tweet.....	29
Tabel 3. 7 Menghitung Euclidean Distance.....	29
Tabel 3. 8 Mengurutkan Jarak .....	30
Tabel 3. 9 Data tweet sesuai kelas .....	31
Tabel 3. 10 Proses Tf idf pada data tweet .....	32
Tabel 3. 11 Hasil probabilitas setiap kata atau token pada setiap kelas .....	35
Tabel 3. 12 hasil preprocessing dok uji coba.....	35
Tabel 4. 1 Data Tweet.....	37
Tabel 4. 2 Case Folding .....	38
Tabel 4. 3 Tokenizing .....	39
Tabel 4. 4 Filtering .....	39
Tabel 4. 5 Stemming.....	40
Tabel 4. 6 Hasil dari klasifikasi textblob .....	43
Tabel 4. 7 Data training dan testing .....	45
Tabel 4. 8 Confusion matrix .....	45
Tabel 4. 9 Hasil Confusion matrix pengujian 1 menggunakan k=3.....	46
Tabel 4. 10 Hasil Confusion matrix pengujian 1 menggunakan k=7.....	47
Tabel 4. 11 Hasil Confusion matrix pengujian 1 menggunakan k=11 .....	48

Tabel 4. 12 Hasil Confusion matrix pengujian 2 menggunakan k=3 .....	48
Tabel 4. 13 Hasil Confusion matrix pengujian 2 menggunakan k=7 .....	49
Tabel 4. 14 Hasil Confusion matrix pengujian 2 menggunakan k=11 .....	49
Tabel 4. 15 Hasil Confusion matrix pengujian 3 menggunakan k=3 .....	50
Tabel 4. 16 Hasil Confusion matrix pengujian 3 menggunakan k=7 .....	50
Tabel 4. 17 Hasil Confusion matrix pengujian 3 menggunakan k=11 .....	51
Tabel 4. 18 Hasil confusion matrix pengujian 1 klasifikasi naïve bayes.....	51
Tabel 4. 19 Hasil confusion matrix pengujian 2 klasifikasi naïve bayes.....	52
Tabel 4. 20 Hasil confusion matrix pengujian 3 klasifikasi naïve bayes.....	52
Tabel 4. 21 Nilai Ketiga Pengujian K-Nearest Neighbors .....	54



## DAFTAR GAMBAR

Gambar 3. 1 Desain Sistem.....	21
Gambar 4. 1 Source Code Textblob Translate.....	42
Gambar 4. 2 Source Code Klasifikasi Data .....	43
Gambar 4. 3 source code TF-IDF .....	44
Gambar 4. 4 Grafik Chart Ketiga Pengujian K-Nearest Neighbors .....	54
Gambar 4. 5 Grafik pengujian Accuracy .....	55
Gambar 4. 6 Grafik pengujian Precision .....	55
Gambar 4. 7 Grafik pengujian Recall .....	56



## INTISARI

Media sosial saat ini merupakan media yang sering digunakan oleh masyarakat Indonesia dalam menyampaikan sebuah pendapat. Salah satu media sosial yang populer di kalangan masyarakat Indonesia saat ini salah satunya adalah Twitter. Dengan meningkatnya jumlah pengguna Twitter dari tahun ke tahun sehingga wirausaha ataupun instansi memanfaatkannya untuk meninjau pendapat mengenai produk jualan. Dengan media sosial Twitter informasi yang didapatkan sangat beragam melalui tweet, tweet sendiri memiliki informasi tertulis yang merupakan data yang dapat diolah menjadi analisis sentimen.

Dalam penelitian ini data yang diperoleh dan diolah adalah data Twitter dengan keyword firstmedia. Data akan dibagi menjadi data latih dan data uji serta dibagi menjadi 3 kelas yaitu positif, netral dan negatif dengan menggunakan metode K-Nearest Neighbor dan Naïve Bayes Classifier.

Berdasarkan hasil pengujian, metode Naïve Bayes Classifier (NBC) mendapatkan akurasi sebesar 72,51%. Dalam penelitian ini klasifikasi data Naïve Bayes Classifier lebih tinggi akurasinya untuk klasifikasi sentimen dibandingkan dengan metode K-Nearest Neighbor yang hanya mendapatkan akurasi sebesar 70,32%.

**Kata kunci:** Analisis Sentimen, Twitter, Naïve Bayes Classifier, K-Nearest Neighbor, FirstMedia.

## ABSTRACT

*Social media is currently a medium that is often used by Indonesian people in expressing an opinion. One of the most popular social media among Indonesian people today is Twitter. With the increasing number of Twitter users from year to year, entrepreneurs or agencies use it to review opinions about their selling products. With the Twitter social media, the information obtained is very diverse through tweets, the tweets themselves have written information which is data that can be used as sentiment analysis.*

*In this study, the data obtained and processed is Twitter data with the keyword firstmedia. The data will be divided into training data and test data and divided into 3 classes namely positive, neutral and negative using the K-Nearest Neighbor and Naïve Bayes Classifier methods.*

*Based on the test results, the Naïve Bayes Classifier (NBC) method obtained an accuracy of 72.51%. In this study, the accuracy of the Naïve Bayes Classifier data classification for sentiment classification is higher than the K-Nearest Neighbor method, which only obtains an accuracy of 70.32%.*

**Keywords:** *Sentiment Analysis, Twitter, Naïve Bayes Classifier, K-Nearest Neighbor, FirstMedia.*