

**SENTIMEN ANALISIS PROVIDER INTERNET XL HOME
PADA TWITTER DENGAN METODE NAÏVE BAYES DAN K-
NEAREST NEIGHBOR**

SKRIPSI

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



disusun oleh

IRFAN KHALIL KHAIRULLAH

17.11.1491

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**SENTIMEN ANALISIS PROVIDER INTERNET XL HOME
PADA TWITTER DENGAN METODE NAÏVE BAYES DAN K-
NEAREST NEIGHBOR**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

IRFAN KHALIL KHAIRULLAH

17.11.1491

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

LEMBAR PERSETUJUAN

SKRIPSI

**SENTIMEN ANALISIS PROVIDER INTERNET XL HOME
PADA TWITTER DENGAN METODE NAÏVE BAYES DAN K-
NEAREST NEIGHBOR**

yang disusun dan diajukan oleh

IRFAN KHALIL KHAIKULLAH

17.11.1491

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 september 2022

Dosen Pembimbing,

Alnul Yaqin, S.Kom., M.Kom

NIK. 190302255

LEMBAR PENGESAHAN

SKRIPSI

**SENTIMEN ANALISIS PROVIDER INTERNET XL HOME
PADA TWITTER DENGAN METODE NAÏVE BAYES DAN K-
NEAREST NEIGHBOR**

yang disusun dan diajukan oleh

IRFAN KHALIL KHAIRULLAH

17.11.1491

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21.september 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Nama dan Gelar Penguji 1

NIK. 190302xxx

Nama dan Gelar Penguji 2

NIK. 190302xxx

Ainul Yaqin, S.Kom., M.Kom

NIK. 190302255

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 September 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.

NIK. 190302096

PERNYATAAN

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : IRFAN KHALIL KHAIRULLAH
NIM : 17.11.1491

Menyatakan bahwa Skripsi dengan judul berikut:

**SENTIMEN ANALISIS PROVIDER INTERNET XL HOME PADA
TWITTER DENGAN METODE NAÏVE BAYES DAN K-NEAREST
NEIGHBOR**

Dosen Pembimbing : Aimul Yaqin, S.Kom., M.Kom

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 rumah 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, 21 September 2022

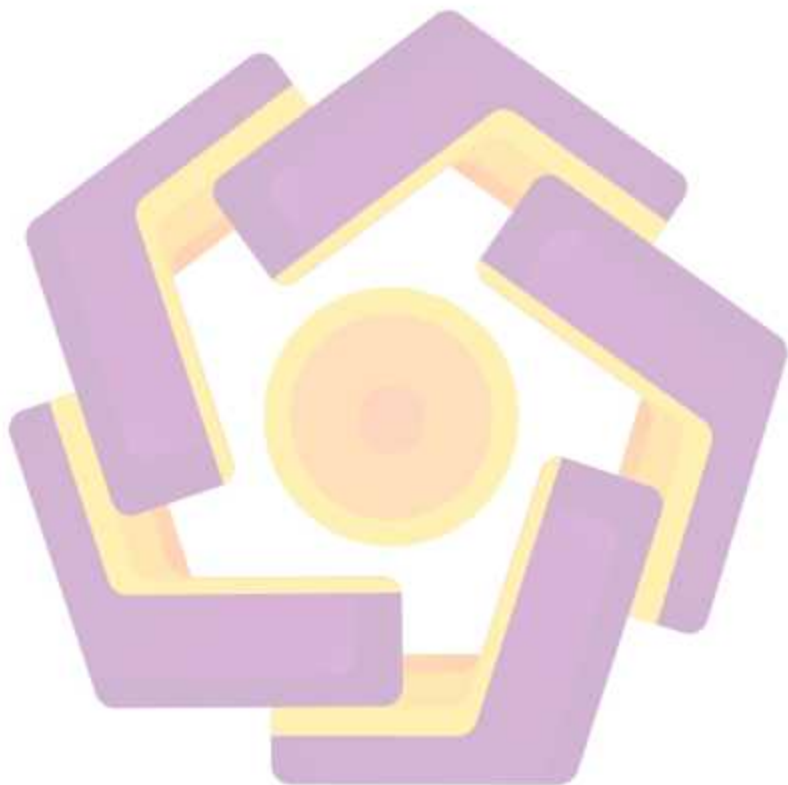
Yang Menyatakan, 21 September 2022



Irfan Khalil Khairullah

MOTTO

"Sebaik baik manusia adalah manusia yang bermanfaat bagi orang lain"



PERSEMBAHAN

Alhamdulillah penulis panjatkan puji-syukur 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, Bapak Aminuddin Dan Ibu Ni Luh Parwati yang selalu mendoakan, memberi semangat serta motivasi supaya dapat menyelesaikan skripsi ini dengan lancar serta bermanfaat bagi semua.
2. Saudariku Zahra Khoirun Nisa yang tidak berhenti untuk memberikan semangat.
3. Bapak. Ainul Yaqin, S.Kom, M.Kom. selaku dosen pembimbing dalam skripsi ini yang tidak lelah untuk tetap membimbing dan mengingatkan penulis dari awal hingga akhir proses pembuatan skripsi.
4. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu baik ilmu akademik maupun ilmu non-akademik selama kuliah.
5. Keluarga besar kelas 17-S11F-09 yang telah bersama-sama menemani selama kuliah. Semoga silaturahmi kita tetap terjaga.
1. 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 kehadirat 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 "SENTIMEN ANALISIS PROVIDER INTERNET XL HOME PADA TWITTER DENGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR".

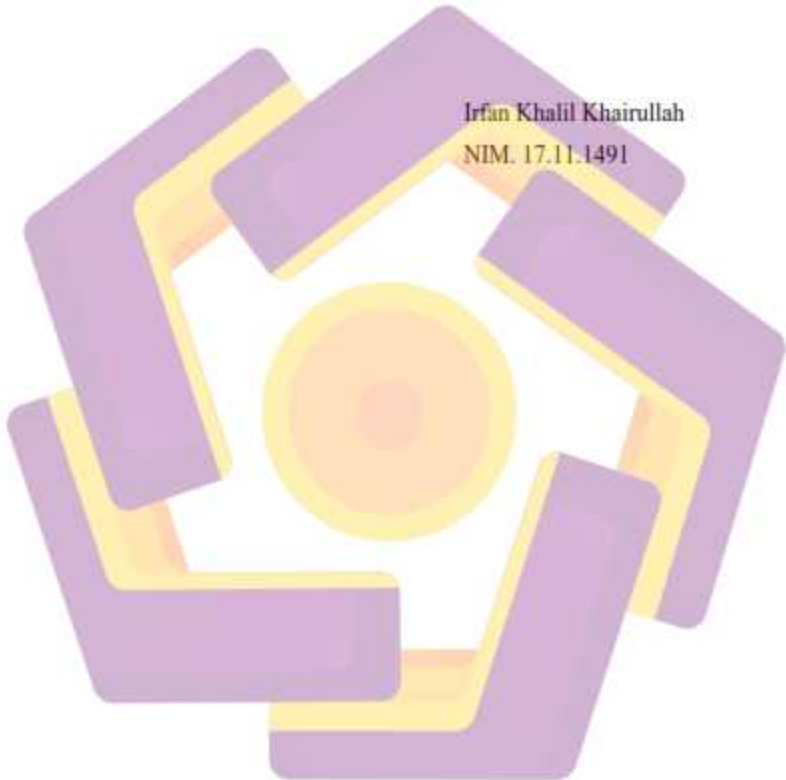
Selama proses pengerjaan skripsi ini penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak dan berkah dari Allah SWT sehingga kendala-kendala tersebut bisa diatasi. Selanjutnya ucapa terima kasih penulis sampaikan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

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak. Ainul Yaqin, S.Kom, M.Kom. 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 (Bpk/Ibu Dosen Penguji, Bpk/Ibu Dosen Penguji, Bpk/Ibu Dosen Penguji) yang telah memberikan masukan terhadap penelitian ini.
6. Kedua orang tua dan keluarga yang selalu memberikan doa, dukungan dan semangat.
7. Keluarga Besar 17-S1IF-09.
8. Serta semua pihak yang tidak bisa penulis sebutkan satu-persatu yang telah membantu dalam penyusunan skripsi ini.

Penulis Menyadari bahwa masih banyak terdapat kekurangan-kekurangan dalam mengerjakan skripsi ini, sehingga penulis mengharapakan adanya saran dan kritik yang membangun demi kesempurnaan skripsi ini.

Yogyakarta, 21 September 2022

Irfan Khalil Khairullah
NIM. 17.11.1491

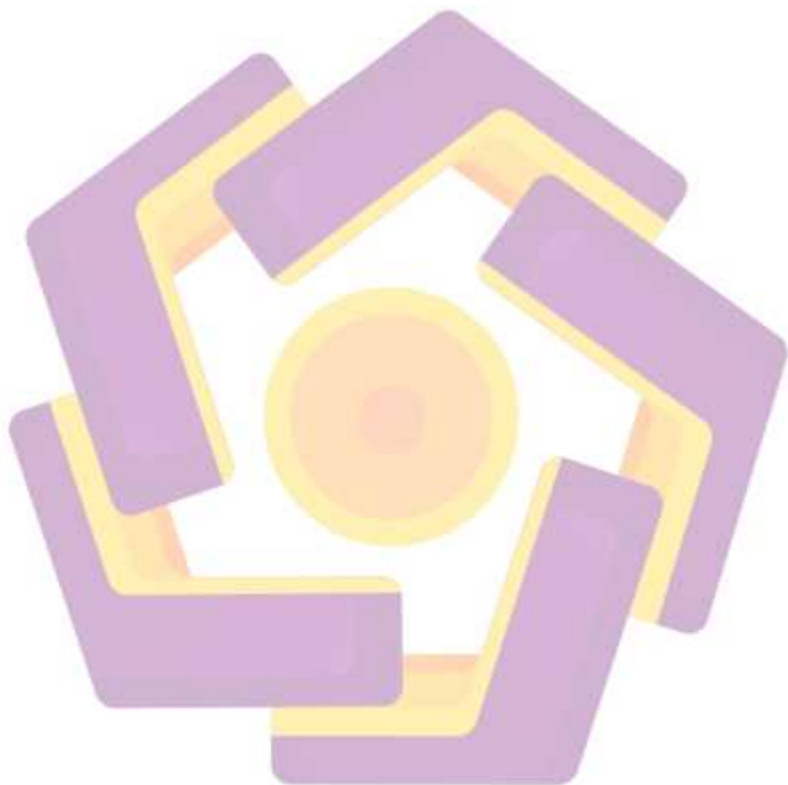


DAFTAR ISI

HALAMAN JUDUL.....	I
LEMBAR PERSETUJUAN.....	II
LEMBAR PENGESAHAN.....	III
PERNYATAAN.....	IV
MOTTO.....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	XII
DAFTAR GAMBAR.....	XIV
INTISARI.....	XV
<i>ABSTRACT</i>	XVI
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	4
1.6 METODE PENELITIAN.....	4
1.7 SISTEMATIKA PENULISAN.....	5
BAB II LANDASAN TEORI.....	7
2.1 TINJAUAN PUSTAKA.....	7
2.1.1 <i>Penelitian Terdahulu</i>	7
2.2 TWITTER.....	12
2.3 ANALISIS SENTIMEN.....	12
2.4 LANDASAN TEORI.....	13

2.4.1	<i>Twitter API</i>	13
2.4.2	<i>Text Mining</i>	14
2.4.3	<i>Text Preprocessing</i>	15
2.4.4	<i>Algoritma Nazief dan Adriani</i>	17
2.4.5	<i>Term Frequency-Inverse Documen Frequency (TF-IDF)</i>	19
2.4.6	<i>K-Nearest Neighbor (KNN)</i>	20
2.4.7	<i>Naïve Bayes Classifier (NBC)</i>	21
BAB III METODE PENELITIAN		23
3.1	DESKRIPSI UMUM	23
3.2	ANALISIS KEBUTUHAN SISTEM	25
3.2.1	<i>Hardware(Perangkat Keras)</i>	25
3.2.2	<i>Software(Perangkat Lunak)</i>	26
3.3	TAHAP PENGUMPULAN DATA	26
3.4	TAHAP PREPROCESSING	27
3.5	PEMBOBOTAN TF-IDF	29
3.6	TAHAP IMPLEMENTASI DAN EVALUASI ALGORITMA NAIVE BAYES CLASIFIER DAN K-NEAREST NEIGHBOR	30
BAB IV HASIL DAN PEMBAHASAN		41
4.1	PROSES PENGUMPULAN DATA	41
4.2	PREPROCESSING DATA	41
4.2.1	<i>Proses Case Folding</i>	42
4.2.2	<i>Proses Tokenizing</i>	42
4.2.3	<i>Proses Filtering</i>	43
4.2.4	<i>Stemming</i>	44
4.3	PEMBOBOTAN TF-IDF	44
4.4	IMPLEMENTASI KLASIFIKASI	45
4.4.1	<i>Klasifikasi Data menggunakan Textblob</i>	45
4.4.2	<i>Pembobotan TF-IDF</i>	47
4.4.3	<i>Uji Coba Sistem</i>	47
4.4.4	<i>Pengujian Naïve Bayes Classifier</i>	49
4.4.5	<i>Pengujian K-Nearest Neighbor</i>	53

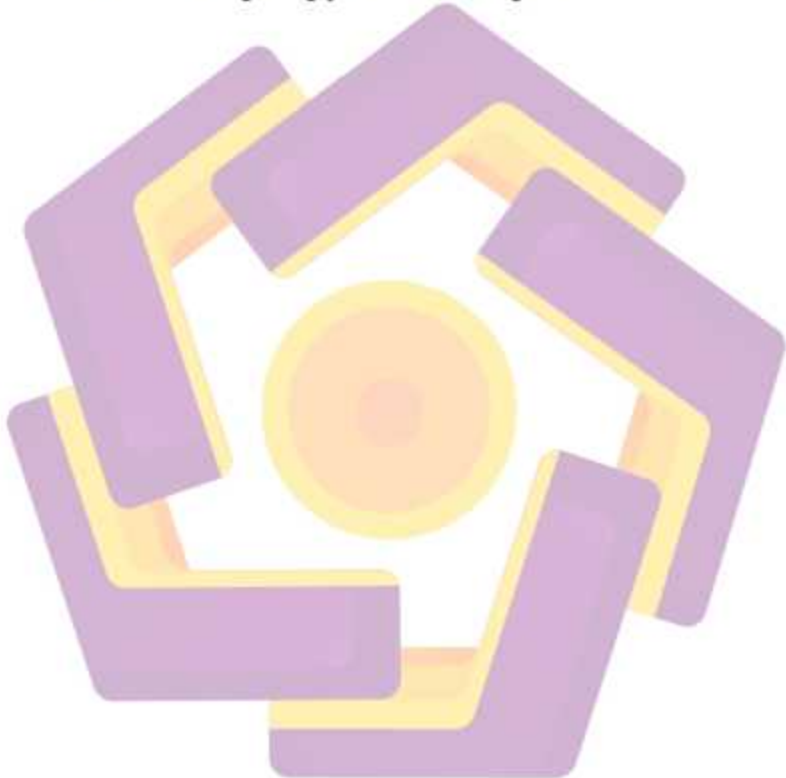
4.4.6 Grafik Hasil Pengujian.....	59
BAB V PENUTUP.....	63
5.1 KESIMPULAN.....	63
5.2 SARAN.....	63
DAFTAR PUSTAKA.....	65



DAFTAR TABEL

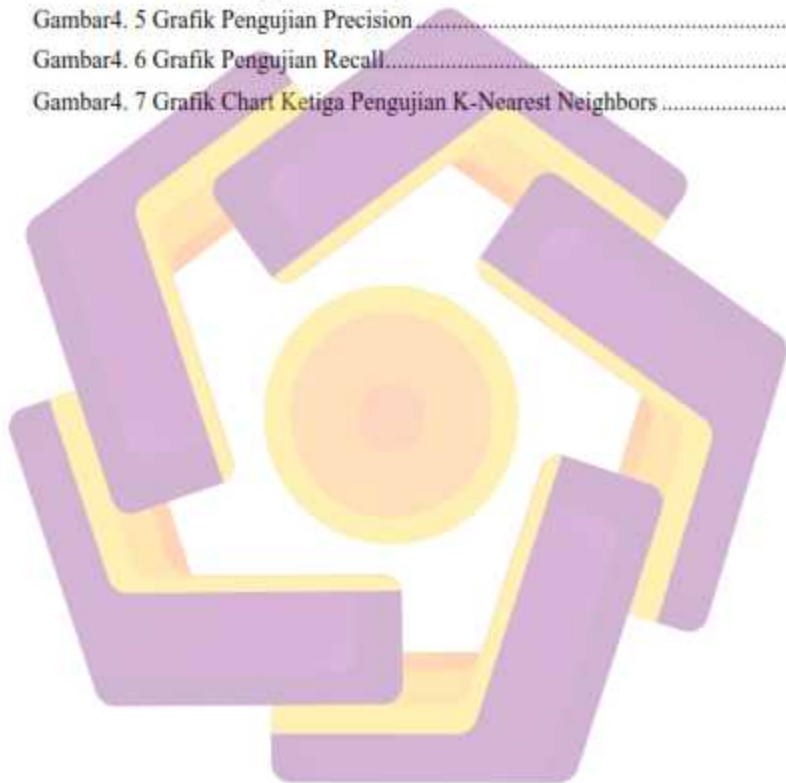
Tabel 2. 1 Studi Literatur Sejenis.....	10
Tabel 3. 1 Proses Case Folding	27
Tabel 3. 2 Proses Tokenizing	28
Tabel 3. 3 Proses Filtering.....	28
Tabel 3. 4 Proses Stemming	29
Tabel 3. 5 Proses TF-IDF	29
Tabel 3. 6 Data tweet.....	31
Tabel 3. 7 Menghitung Euclidean Distance	32
Tabel 3. 8 Mengurutkan Jarak	32
Tabel 3. 9 data tweet sesuai kelas.....	33
Tabel 3. 10 Proses Tf.idf pada data tweet.....	34
Tabel 3. 11 Hasil probabilitas setiap kata atau token pada setiap kelas	39
Tabel 3. 12 table hasil preprocessing dok ujicoba.....	39
Tabel 4. 1 Data Tweet.....	41
Tabel 4. 2 Case Folding.....	42
Tabel 4. 3 Tokenizing.....	42
Tabel 4. 4 Filtering	43
Tabel 4. 5 Stemming.....	44
Tabel 4. 6 Hasil dari klasifikasi textblob.....	46
Tabel 4. 7 Data training dan testing.....	47
Tabel 4. 8 confusion matrix.....	48
Tabel 4. 9 Hasil confusion matrix pengujian 1 klasifikasi naïve bayes	49
Tabel 4. 10 Hasil confusion matrix pengujian 2 klasifikasi naïve bayes	50
Tabel 4. 11 Hasil confusion matrix pengujian 3 klasifikasi naïve bayes	51
Tabel 4. 12 Hasil Confusion matrix pengujian 1 menggunakan k=3	53
Tabel 4. 13 Hasil Confusion matrix pengujian 1 menggunakan k=7	53
Tabel 4. 14 Hasil Confusion matrix pengujian 1 menggunakan k=11	54
Tabel 4. 15 Hasil nilai rata rata pengujian 1	54
Tabel 4. 16 Hasil Confusion matrix pengujian 2 menggunakan k=3	55
Tabel 4. 17 Hasil Confusion matrix pengujian 2 menggunakan k=7	55

Tabel 4. 18 Hasil Confusion matrix pengujian 2 menggunakan $k=11$	56
Tabel 4. 19 Hasil nilai rata rata pengujian 2.....	56
Tabel 4. 20 Hasil Confusion matrix pengujian 3 menggunakan $k=3$	57
Tabel 4. 21 Hasil Confusion matrix pengujian 3 menggunakan $k=7$	57
Tabel 4. 22 Hasil Confusion matrix pengujian 3 menggunakan $k=11$	58
Tabel 4. 23 Hasil nilai rata rata pengujian 3.....	58
Tabel 4. 24 Nilai Ketiga Pengujian K-Nearest Neighbors	62



DAFTAR GAMBAR

Gambar 3. 1 Desain Sistem	25
Gambar4. 1 Source Code Translate.....	45
Gambar4. 2 Source Code Klasifikasi Data.....	46
Gambar4. 3 source code TF-IDF.....	47
Gambar4. 4 Grafik Pengujian Accuracy.....	59
Gambar4. 5 Grafik Pengujian Precision.....	60
Gambar4. 6 Grafik Pengujian Recall.....	61
Gambar4. 7 Grafik Chart Ketiga Pengujian K-Nearest Neighbors	62



INTISARI

Media sosial saat ini merupakan media sosial yang sering digunakan oleh kalangan 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 jualnya. Dengan media sosial Twitter informasi yang didapatkan sangat beragam melalui tweet, dari informasi yang ditulis melalui tweet terdapat data yang bisa diolah menjadi analisis sentimen.

Dalam penelitian ini data yang diperoleh dan di gunakan adalah data Twitter dengan keyword *xlhome*. Data akan di bagi menjadi data latih dan data uji serta dibagi menjadi 3 kelas yaitu *positive*, *netral* dan *negative* dengan menggunakan metode *Naive Bayes Classifier* dan *K-Nearest Neighbor*.

Berdasarkan hasil pengujian, metode *Naive Bayes Classifier* (NBC) mendapatkan akurasi rata-rata mencapai 68%. Dalam penelitian ini klasifikasi data NBC lebih tinggi akurasinya untuk klasifikasi sentimen dibandingkan dengan metode KNN yang hanya mencapai akurasi tertinggi mencapai 64%.

Kata Kunci: *Analisis Sentimen, Twitter, Naive Bayes Classifier, K-Nearest Neighbor, XLHome*

ABSTRACT

Social media is currently a social media that is often used by Indonesian people to convey 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 social media Twitter, the information obtained is very diverse through tweets, from information written via tweets, there is data that can be processed into sentiment analysis.

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

*Based on the test results, the *Naïve Bayes Classifier (NBC)* method obtains an average accuracy of 68%. In this study, the classification of NBC data has a higher accuracy for sentiment classification compared to the *KNN* method, which only achieves the highest accuracy of 64%.*

Keywords: *Sentiment Analysis, Twitter, Naive Bayes Classifier, K-Nearest Neighbor, XLHome*