

**IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR*
UNTUK SENTIMEN KEPUASAN PELANGGAN
PADA AKUN TWITTER INDIHOME**

SKRIPSI



disusun oleh

Dyah Wasistha

17.11.1527

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR*
UNTUK SENTIMEN KEPUASAN PELANGGAN
PADA AKUN TWITTER INDIHOME**

Skripsi

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Dyah Wasistha
17.11.1527

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR*
UNTUK SENTIMEN KEPUASAN PELANGGAN
PADA AKUN TWITTER INDIHOME**

yang dipersiapkan dan disusun oleh

Dyah Wasistha

17.11.1527

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 April 2020

Dosen Pembimbing,

Sumarni Adi, S.Kom, M.Cs.

NIK. 190302256

PENGESAHAN
SKRIPSI
IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR*
UNTUK SENTIMEN KEPUASAN PELANGGAN
PADA AKUN TWITTER INDIHOME

yang dipersiapkan dan disusun oleh

Dyah Wasistha

17.11.1527

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 April 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Supriatin, M.Kom
NIK. 190302239

Nuraini, M.Kom
NIK. 190302066

Sumarni Adi, S.Kom, M.Cs
NIK. 190302256

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 April 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 21 April 2021



Dyah Wasistha

17.11.1527

MOTTO

” Dan bahwasanya seorang manusia tiada memperoleh selain apa yang telah diusahakannya”
(An Najm : 39)

”Success is the ability to go from one failure to another with no loss of enthusiasm”
(Sir Winston Churchill)



PERSEMBAHAN

Puji syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah sehingga penulis masih diberikan kesempatan untuk menyelesaikan skripsi ini sebagai salah satu syarat untuk mendapatkan gelar sarjana. Walaupun jauh dari kata sempurna, penulis sangat bersyukur telah mencapai titik ini yaitu dapat menyelesaikan skripsi ini di waktu yang tepat.

Penulis ingin mengucapkan banyak terima kasih dan mempersembahkan skripsi ini kepada pihak yang telah membantu dalam penyusunan skripsi ini.

1. Kepada orang tua saya, Bapak Slamet Nurwanto dan Ibu Sumarni. Karena kalian berdua, hidup terasa begitu mudah dan penuh kebahagiaan. Terima kasih atas segala bentuk dukungan, semangat, doa yang telah dipanjatkan selama saya menempuh pendidikan dan selalu menjaga saya dalam mengejar impian saya apa pun itu.
2. Kepada ibu Sumarni Adi, S.Kom, M.Cs selaku dosen pembimbing saya dalam penyusunan skripsi ini. Terima kasih atas kesabaran dan waktunya untuk membimbing saya sehingga saya bisa menyelesaikan skripsi ini dengan baik dan tepat waktu.
3. Kepada kakak saya tercinta Ratri Octaviani untuk segala doa, dukungan dan motivasinya untuk saya menyelesaikan skripsi ini.
4. Kepada teman-teman kelas 17 Informatika 09 yang menemani saya selama menempuh pendidikan di sini.
5. Kepada teman-teman yang telah membantu khususnya Sis Romlan untuk segala support dan telah rela mendengarkan keluh kesah saya selama

perkuliahan maupun pengerjaan skripsi ini. Terimakasih telah menyediakan telinga untuk mendengarkan.



KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah, serta kekuatan sehingga saya dapat menyelesaikan skripsi saya yang berjudul Implementasi Algoritma *K-Nearest Neighbor* untuk Sentimen Kepuasan Pelanggan pada Akun Twitter Indihome.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi informatika fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku rektor Universitas Amikom Yogyakarta
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Ibu Sumarni Adi, S.Kom, M.Cs, selaku dosen pembimbing yang telah memberikan saran, arahan, motivasi dan bimbingannya untuk saya dalam pembuatan skripsi ini
4. Kedua orang tua serta kakak tercinta atas dukungan, semangat, doa yang selalu dipanjatkan untuk saya sehingga bisa menyelesaikan skripsi ini dengan tepat waktu

5. Teman seperjuangan 17 Informatika 09 yang menjadi teman selama proses perkuliahan
6. Teman terdekatku yang selalu memberikan semangat dan masukan dalam pembuatan skripsi ini
7. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung

Semoga Allah senantiasa membalas semua kebaikan untuk semua pihak yang membantu saya dalam menyelesaikan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 21 April 2021

Dyah Wasistha

DAFTAR ISI

Contents

PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN	iv
MOTTO	v
PERSEMBAHAN	vi
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metode Penelitian	4
1.6.1 Metode Pengumpulan Data	4
1.6.2 Metode Analisis	5
1.6.3 Metode Implementasi	7
1.6.4 Metode Testing	7
1.7 Sistematika Penelitian	7
BAB II	9
LANDASAN TEORI	9
2.1 Kajian Pustaka	9
2.2 Dasar Teori	13
2.2.1 Social Media	13
2.2.2 Text Mining	14

2.2.3 Sentimen Analisis	15
2.2.4 Kelas Sentimen	16
2.2.5 Data <i>Crawling</i>	17
2.2.6 <i>Preprocessing</i>	18
2.2.7 <i>K-Nearest Neighbor</i>	20
2.2.8 <i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	21
2.2.9 Confusion Matrix	22
BAB III	25
ANALISIS DAN PERANCANGAN	25
3.1 Tinjauan Umum	25
3.2 Alat dan Bahan Penelitian	26
3.2.1 Perangkat Keras (<i>Hardware</i>)	26
3.2.2 Perangkat Lunak (<i>Software</i>)	26
3.3 Alur Penelitian	27
3.3.1 <i>Crawling</i> Data Twitter	28
3.3.2 <i>Preprocessing Dataset</i>	32
3.3.3 Data Split	40
3.3.4 <i>Feature Extraction</i>	41
3.3.5 <i>K-Nearest Neighbor Classification</i>	49
3.3.6 Confusion Matriks	61
3.4 Perancangan Antar Muka (<i>Interface</i>)	61
3.4.1 <i>Interface</i> Halaman Utama	61
3.4.2 <i>Interface</i> Upload Data	62
3.4.3 <i>Interface Preprocessing</i>	63
3.4.4 <i>Interface</i> TF-IDF	64
3.4.5 <i>Interface</i> Klasifikasi KNN	65
3.4.6 <i>Interface</i> Tes Model	66
BAB IV	68
IMPLEMENTASI DAN PEMBAHASAN	68
4.1 Deskripsi Implementasi	68
4.2 Implementasi Pengumpulan Data	68
4.3 Implementasi Pelabelan Data	70
4.4 Implementasi Perancangan Fungsi dan Sistem	70



4.4.1 Implementasi Fungsi <i>Preprocessing</i> Data	71
4.4.2 Implementasi Data Split	73
4.4.3 Implementasi TF-IDF	73
4.4.4 Implementasi Algoritma <i>K-Nearest Neighbors</i>	73
4.4.5 Implementasi Confusion Matrix.....	74
4.5 Implementasi Interface	74
4.5.1 Halaman Utama	75
4.5.2 Halaman Upload Data	75
4.5.3 Halaman <i>Preprocessing</i>	76
4.5.4 Halaman TF-IDF	77
4.5.5 Halaman Klasifikasi KNN	78
4.5.6 Halaman Tes Model	79
BAB V.....	80
PENUTUP.....	80
5.1 Kesimpulan.....	80
5.2 Saran	80
DAFTAR PUSTAKA	82

DAFTAR TABEL

Tabel 2. 1 Keaslian Peneliti	10
Tabel 2. 2 Rumus Confesion Matrix	23
Tabel 3. 1 Spesifikasi Hardware	26
Tabel 3. 2 Perangkat Lunak (Software)	27
Tabel 3. 3 Dataset Utuh.....	29
Tabel 3. 4 Contoh Data Tweet	30
Tabel 3. 5 Labeling Data Tweet.....	31
Tabel 3. 6 Contoh Username Removal	32
Tabel 3. 7 Contoh <i>Case folding</i>	33
Tabel 3. 8 Contoh Punctuation Removal	34
Tabel 3. 9 Contoh Cleaning	36
Tabel 3. 10 Contoh Tokenizing.....	37
Tabel 3. 11 Contoh Stemming	38
Tabel 3. 12 Contoh Stopword Removal.....	39
Tabel 3. 13 Term Kata	41
Tabel 3. 14 Perhitungan IDF.....	43
Tabel 3. 15 Perhitungan TF.IDF	45
Tabel 3. 16 Data Tweet.....	50
Tabel 3. 17 Data <i>Preprocessing</i>	51
Tabel 3. 18 Perhitungan TF-IDF.....	53
Tabel 3. 19 Perhitungan Panjang Vector.....	56
Tabel 3. 20 Kemiripan Hasil	60
Tabel 3. 21 Hasil Kemiripan Terdekat.....	60
Tabel 4. 1 Hasil Klasifikasi KNN	78

DAFTAR GAMBAR

Gambar 2. 1 <i>Crawling</i>	18
Gambar 3. 1 Alur Penelitian.....	28
Gambar 3. 2 Data Split.....	40
Gambar 3. 3 Flowchart KNN	49
Gambar 3. 4 Interface Utama	62
Gambar 3. 5 Interface Upload Data	63
Gambar 3. 6 Interface <i>Preprocessing</i>	64
Gambar 3. 7 Interface TF-IDF	65
Gambar 3. 8 Interface Klasifikasi KNN.....	66
Gambar 3. 9 Interface Tes Model	67
Gambar 4. 1 Script <i>Crawling</i> Data Twitter.....	69
Gambar 4. 2 Dataset Utuh.....	69
Gambar 4. 3 Script Simpan Data.CSV	69
Gambar 4. 4 Label Data	70
Gambar 4. 5 Script Open File.CSV.....	71
Gambar 4. 6 Script <i>Case folding</i>	71
Gambar 4. 7 Script Remove Punctuation.....	72
Gambar 4. 8 Script Stopword Removal	72
Gambar 4. 9 Script Stemming.....	73
Gambar 4. 10 Implementasi Data Split.....	73
Gambar 4. 11 Implementasi TF-IDF.....	73
Gambar 4. 12 Implementasi Algoritma KNN	74
Gambar 4. 13 Implementasi Confusion Matrix.....	74
Gambar 4. 14 Implementasi Halaman Utama	75
Gambar 4. 15 Implementasi Upload Data.....	76
Gambar 4. 16 Implementasi <i>Preprocessing</i>	77
Gambar 4. 17 Implementasi TF-IDF.....	77
Gambar 4. 18 Implementasi Klasifikasi KNN	78
Gambar 4. 19 Implementasi Tes Model.....	79

INTISARI

Perkembangan sosial media di Indonesia cukup pesat, hal ini dibuktikan dengan meningkatnya penggunaan Twitter sebagai salah satu media marketing bagi sebuah perusahaan. Dengan adanya hal ini, tentunya dimanfaatkan oleh penyedia jasa internet Indihome yang saat ini penggunaannya semakin meningkat. Oleh sebab itu, seiring dengan peningkatan jumlah pengguna provider Indihome, maka dapat memunculkan berbagai opini khususnya pada platform twitter, baik yang bersifat pujian maupun keluhan. Dengan munculnya berbagai opini tersebut, peneliti melakukan analisis sentimen untuk mengetahui apakah pelanggan puas atau tidak terhadap penyedia layanan internet Indihome. Banyaknya opini yang dituliskan di twitter akan dijadikan sebagai dataset yang berjumlah 1500 data tweet. Dataset tersebut akan diklasifikasikan apakah dataset tersebut cenderung beropini positif atau negatif.

Adapun beberapa tahapan yang dilakukan yaitu pengumpulan data atau *crawling dataset* dengan menggunakan API twitter, *preprocessing* untuk menghilangkan *noise* yang tidak diperlukan, *feature extraction* dengan metode TF-IDF, klasifikasi dengan Algoritma *K-Nearest Neighbor* dan dievaluasi dengan menggunakan metode *Confusion Matrix*.

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan algoritma KNN, didapatkan tingkat akurasi sebesar 82,6%, *precision score* 83.17%, *f1 score* 87.2% dan *recall score* sebesar 91.75%.

Kata Kunci: KNN, Analisis Sentimen, Twitter, Indihome

ABSTRACT

The development of social media in Indonesia is quite fast, this is evidenced by the increasing number of Twitter users as marketing media for a company. It is utilized by internet service provider Indihome, whose users are currently increasing. Therefore, along with the increase in the number of users of Indihome providers, various opinions especially on the Twitter platform can be raised, both in the form of praise and complaints. With the emergence of these various opinions, researchers conducted a sentiment analysis to determine whether or not customers were satisfied with the Indihome internet service provider. The number of opinions written on Twitter will serve as a dataset of 1500 tweets. The dataset will be classified whether it tends to have a positive or negative opinion.

The steps taken are data collection using the Twitter API, preprocessing to remove unnecessary noise, feature extraction using the TF-IDF method, classification using the K-Nearest Neighbor algorithm and evaluating using the Confusion Matrix method.

Based on the results of the research that has been done, the accuracy rate is 82.6%, precision score 83.17%, f1 score 87.2% and recall score 91.75%.

Keywords: *KNN, Analisis Sentimen, Twitter, Indihome*