

**PERBANDINGAN ALGORITMA K – NEAREST NEIGHBOR DAN  
NAÏVE BAYES CLASSIFIER DALAM KLASIFIKASI KARAKTER  
INDIVIDU PADA LINIMASA AKUN TWITTER**

**SKRIPSI**



disusun oleh

**Aminudin Noor Ichsan**

**16.11.0039**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**PERBANDINGAN ALGORITMA K – NEAREST NEIGHBOR DAN  
NAÏVE BAYES CLASSIFIER DALAM KLASIFIKASI KARAKTER  
INDIVIDU PADA LINIMASA AKUN TWITTER**

**SKRIPSI**

untuk memenuhi sebagian persyaratan mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Aminudin Noor Ichsan**

**16.11.0039**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA K – NEAREST NEIGHBOR DAN  
NAÏVE BAYES CLASSIFIER DALAM KLASIFIKASI KARAKTER  
INDIVIDU PADA LINIMASA AKUN TWITTER**


yang dipersiapkan dan disusun oleh

**Aminudin Noor Ichsan**

**16.11.0039**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 November 2019

Dosen Pembimbing,

  
**Prof. Dr. Ema Utami, S.Si, M.Kom.**  
**NIK. 190302037**

**PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA K – NEAREST NEIGHBOR DAN  
NAÏVE BAYES CLASSIFIER DALAM KLASIFIKASI KARAKTER  
INDIVIDU PADA LINIMASA AKUN TWITTER**

yang dipersiapkan dan disusun oleh

**Aminudin Noor Ichsan**

**16.11.0039**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 13 November 2019

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Sumarni Adi, S.Kom, M.Cs**

**NIK. 190302256**

**Anggit Dwi Hartanto, M.Kom**

**NIK. 190302163**

**Arifiyanto Hadinegoro, S.Kom, M.T**

**NIK. 190302289**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 November 2019

**DEKAN FAKULTAS ILMU KOMPUTER**



**Krisnawati, S.Si, M.T**

**NIK. 190302038**

## PERNYATAAN

Saya yang bertandatangan dibawah 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 tanggungjawab saya pribadi.

Yogyakarta, 17 November 2019



Aminudin Noor Ichsan  
NIM. 16.11.0039

## MOTTO

*“Learn from yesterday, live for today, hope for tomorrow. The important thing is not stop questioning.”*  
(Albert Einstein)

*“I have not failed. I’ve just found 10.000 ways that won’t work.”*  
(Thomas Edison)

*“It does not matter how slowly you go, so long as you do not stop.”*  
(Confucius)

*“The biggest risk is not taking a risk. In a world that’s changing really quickly, the only strategy that is guaranteed to fail is not taking risk.”*  
(Mark Zuckerberg)

*“Jika sekarang kamu tidak mulai berjalan, maka besok kamu harus berlari”*  
(Aji Kamaludin)





## PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah kepada Allah SWT, akhirnya skripsi ini telah berhasil diselesaikan penulis dengan baik dan sesuai target. Tidaklah benar, jika penulis mengatakan didalam skripsi ini tidak ada bantuan serta dukungan dari orang lain. Maka dari itu, pada lembar ini izinkan penulis mengucapkan rasa terimakasih dan sedikit mencurahkan perasaannya kepada :

1. Kedua orang tua tercinta, Bapak Tamino dan Ibu Yami Badilah. Terima kasih karena telah bersusah payah membanting tulang untuk membiayai kuliah serta hidup penulis. Semoga dengan diselesaikannya skripsi ini, menjadikan suatu batu loncatan bagi penulis untuk bisa menjadi salah satu orang yang beruntung dan sukses sehingga dapat membahagiakan kalian. Amin..
2. Saudara dan Saudari ku tersayang, Aminullah Noor Adnan, Aminah Noor Aisyah, dan Aminah Noor Salamah yang sudah memberikan dukungan dan doa kepada penulis.
3. Teman-teman sekelas, 16-S1 IF-01 yang mohon maaf tidak bisa penulis sebutkan satu persatu karena terlalu banyak.
4. Sahabat pejuang skripsi kontrakan Bos Pared dan kontrakan Gelatik : Farid, Andi, Hananto, Ariel, Taufiq, Arik, Arif Marda, Arif Setyo, Indra, Ardi Bayu, Dian Noviardi, Wahyudi, dan Irfan. Terima kasih untuk dukungan, doa, tempat, dan waktu yang sudah kita habiskan bersama. Semoga kita semua menjadi orang-orang yang beruntung.
5. Sahabat seperjuangan, Aji Kamaludin, Harish Setyo Hudnanto, dan Maful Prayoga Arnandi. Terimakasih sudah berjuang bersama, saling bertukar pemikiran, dan berkeluh kesah bersama. Semoga kita tidak pernah melupakan kenangan perjuangan ini.
6. Pihak Gojek dan juga Grab yang sudah memberikan berbagai macam promo yang membantu penulis dalam mengerjakan Skripsi.

## KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT atas segala limpahan ridho, hidayah, serta inayah-Nya sehingga penelitian dengan judul “Perbandingan Algoritma K – Nearest Neighbor dan Naïve Bayes Classifier Dalam Klasifikasi Karakter Individu Pada Linimasa Akun Twitter” ini dapat penulis selesaikan dengan baik dan lancar. Skripsi ini disusun untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer Universitas Amikom Yogyakarta. Dengan segala keterbatasan yang penulis miliki, masih banyak kekurangan-kekurangan yang harus diperbaiki. Semoga hasil penelitian ini dapat berguna, khususnya bagi dunia pendidikan.

Dalam penulisan Skripsi ini, penulis banyak mendapat bantuan dari berbagai pihak. Oleh karena itu, ucapan terima kasih penulis sampaikan kepada :

- Kedua Orang Tua Penulis.
- Kepada Dosen Pembimbing Prof. Dr. Ema Utami, S.Si., M.Kom, Bapak Anggit Dwi Hartanto, M.Kom, dan Ibu Sumarni Adi, S.Kom., M.Cs.
- Kepada sahabat dan teman-teman yang telah memberikan dukungan serta doa yang tulus kepada penulis.
- Kepada kampus tercinta, Universitas Amikom Yogyakarta.

Yogyakarta, 17 November 2019

Aminudin Noor Ichsan



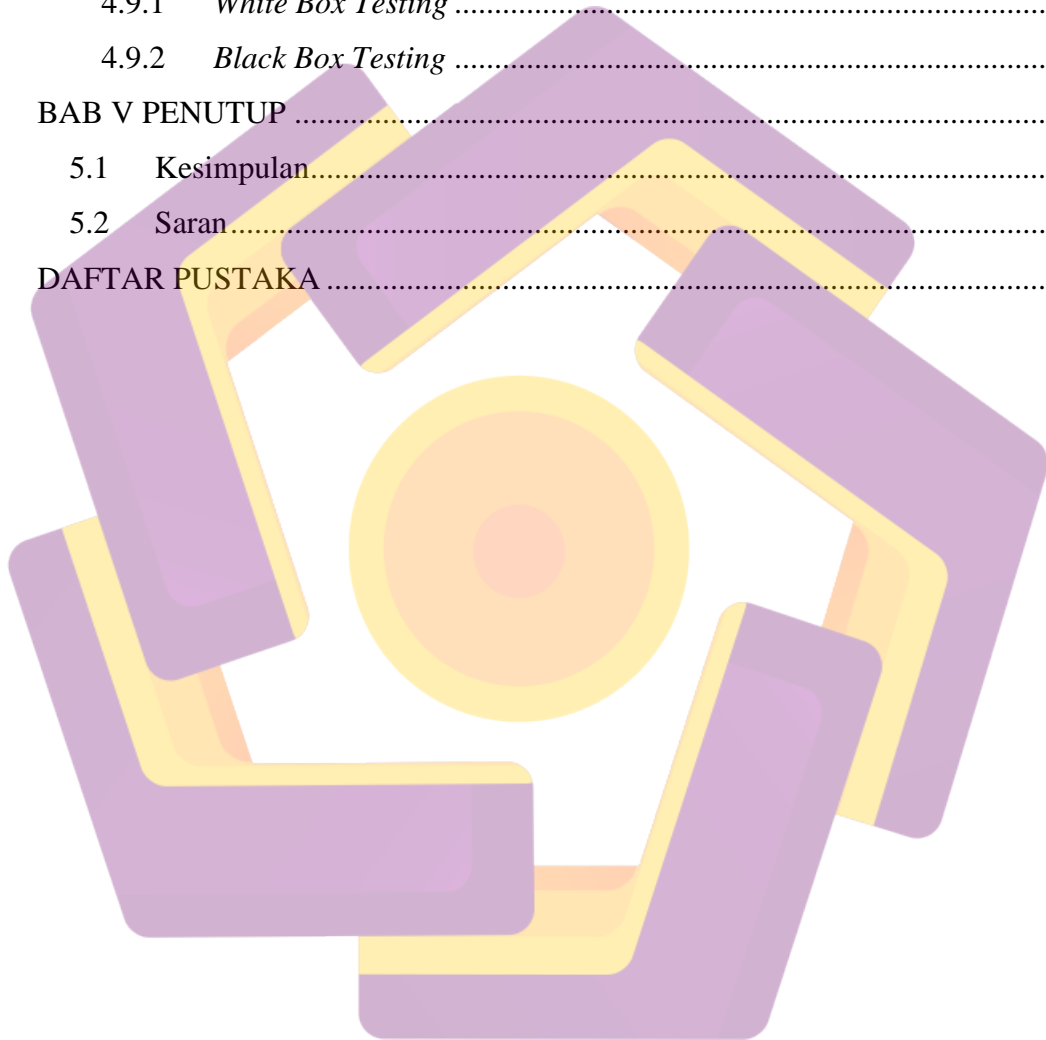
## DAFTAR ISI

JUDUL .....	I
PERSETUJUAN .....	II
PENGESAHAN .....	III
PERNYATAAN .....	IV
MOTTO .....	V
PERSEMBAHAN .....	VI
KATA PENGANTAR .....	VII
DAFTAR ISI .....	VIII
DAFTAR TABEL .....	XII
DAFTAR GAMBAR .....	XIV
DAFTAR LAMPIRAN .....	XVI
INTISARI .....	XVII
<i>ABSTRACT</i> .....	XVIII
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Maksud dan Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	5
1.6 Metodologi Penelitian .....	5
1.6.1 Metode Pengumpulan Data .....	5
1.6.2 Metode Analisis .....	6
1.6.3 Metode Perancangan .....	6
1.6.4 Implementasi .....	6
1.6.5 Pengujian .....	6
1.7 Sistematika Penulisan .....	6
<b>BAB II LANDASAN TEORI</b> .....	8
2.1 Tinjauan Pustaka .....	8
2.2 Dasar Teori .....	10

2.2.1	Pengertian <i>Twitter</i> .....	10
2.2.2	Pengertian Karakter.....	11
2.2.3	Teori DISC ( <i>Dominance, Influence, Steadiness, dan Compliance</i> ).11	
2.2.4	Data Mining .....	13
2.2.4.1	Tahap – Tahap Data Mining .....	14
2.2.4.2	Pengelompokan Data Mining .....	15
2.2.5	<i>Scraping</i> .....	17
2.2.6	Tokenisasi .....	18
2.2.7	Pembobotan TF-IDF .....	18
2.2.8	Algoritma <i>K – Nearest Neighbor</i> .....	19
2.2.9	<i>Euclidean Distance</i> .....	20
2.2.10	Algoritma <i>Naïve Bayes Classifier</i> .....	20
2.2.11	Validasi dan Evaluasi .....	21
<b>BAB III ANALISIS DAN PERANCANGAN .....</b>		<b>24</b>
3.1	Analisis Sistem.....	24
3.1.1	Analisis Kebutuhan Fungsional .....	24
3.1.2	Analisis Kebutuhan Non-Fungsional.....	24
3.2	Gambaran Umum Sistem .....	25
3.3	Analisis Algoritma .....	26
3.3.1	Pengumpulan Data .....	26
3.3.2	Validasi Psikolog .....	27
3.3.3	<i>Preprocessing</i> .....	28
3.3.3.1	<i>Case Folding</i> .....	29
3.3.3.2	<i>Tweet Cleaning</i> .....	30
3.3.3.3	Tokenizing .....	32
3.3.4	Pembobotan TF-IDF .....	33
3.3.4.1	Menghitung TF Pada <i>Tweet</i> .....	34
3.3.4.2	Menghitung DF Kata Pada Semua <i>Tweet</i> .....	35
3.3.4.3	Menghitung Nilai TF-IDF .....	35
3.3.4.4	Pengelompokan TF-IDF Berdasarkan Label .....	36
3.3.5	<i>Naïve Bayes Classifier</i> .....	37

3.3.6	<i>K – Nearest Neighbor</i> .....	38
3.3.7	<i>Confusion Matrix</i> .....	40
3.3.7.1	Validasi Data Set .....	40
3.3.7.2	Hasil Olah Data Menggunakan Algoritma <i>Naïve Bayes Classifier</i> .....	42
3.3.7.3	Hasil Olah Data Menggunakan Algoritma <i>K – Nearest Neighbor</i> .....	43
3.3.7.4	Hasil Evaluasi .....	43
3.4	Analisis Kebutuhan Data.....	46
3.4.1	Perancangan Data Set.....	46
3.4.2	Perancangan Data Training.....	47
3.5	Perancangan Antarmuka .....	47
3.5.1	Perancangan Halaman <i>Scraping</i> .....	47
3.5.2	Perancangan Halaman Detail Akun .....	49
3.5.3	Perancangan Halaman <i>Preprocessing</i> .....	49
3.5.4	Perancangan Halaman Detail <i>Preprocessing</i> .....	50
3.5.5	Perancangan Halaman Klasifikasi .....	51
3.5.6	Perancangan Halaman Detail Klasifikasi.....	52
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN</b> .....		54
4.1	Deskripsi Implementasi.....	54
4.2	Implementasi Pengumpulan Data <i>Tweet</i> .....	54
4.3	Implementasi Pelabelan Dataset Melalui Validasi Psikolog.....	58
4.4	Implementasi <i>Preprocessing</i> .....	60
4.5	Implementasi Pembobotan dan Klasifikasi .....	62
4.6	Evaluasi <i>Confusion Matrix</i> .....	67
4.6.1	Hasil Olah Data Menggunakan Algoritma <i>Naïve Bayes Classifier</i> .....	67
4.6.2	Hasil Olah Data Menggunakan Algoritma <i>K – Nearest Neighbor</i> .....	67
4.6.3	Hasil Evaluasi .....	68
4.7	Evaluasi <i>3 – Fold Cross Validation</i> .....	75
4.7.1	Hasil Olah Data Menggunakan Algoritma <i>Naïve Bayes Classifier</i> .....	75
4.7.2	Hasil Olah Data Menggunakan Algoritma <i>K – Nearest Neighbor</i> .....	75
4.6.3	Hasil Evaluasi .....	76

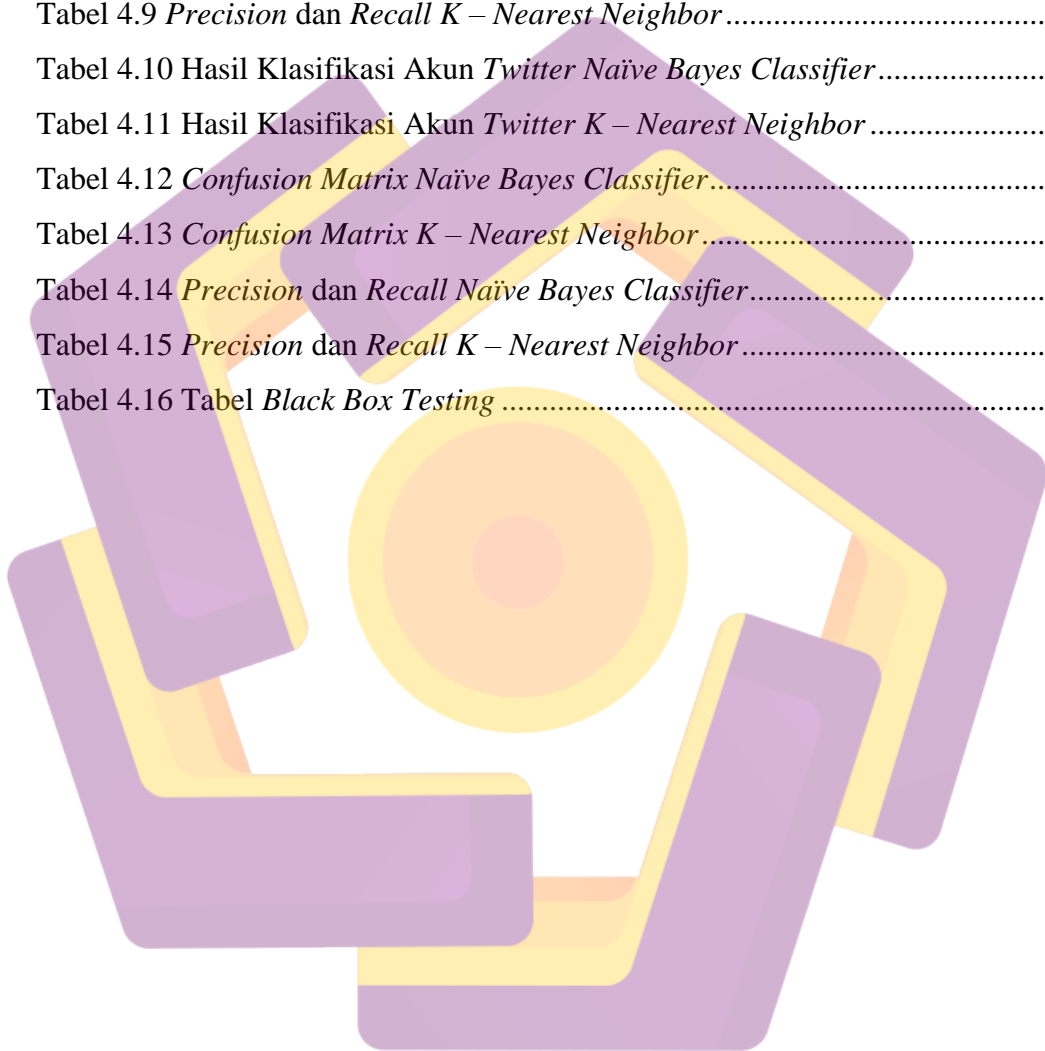
4.8	Implementasi Antarmuka .....	81
4.8.1	Implementasi Antarmuka Halaman Akun <i>Twitter</i> .....	81
4.8.2	Implementasi Antarmuka Halaman <i>Preprocessing</i> .....	82
4.8.3	Implementasi Antarmuka Halaman Klasifikasi .....	84
4.9	Pengujian Sistem .....	87
4.9.1	<i>White Box Testing</i> .....	87
4.9.2	<i>Black Box Testing</i> .....	88
BAB V PENUTUP .....		89
5.1	Kesimpulan.....	89
5.2	Saran.....	90
DAFTAR PUSTAKA .....		92



## DAFTAR TABEL

Tabel 2.1 Perbandingan Pustaka .....	9
Tabel 2.2 <i>Confusion Matrix</i> .....	22
Tabel 3.1 Contoh Data <i>Tweet</i> .....	26
Tabel 3.2 Contoh Pelabelan Kata.....	28
Tabel 3.3 Contoh Pelabelan Akun .....	28
Tabel 3.4 <i>Case Folding</i> .....	29
Tabel 3.5 <i>Tweet Cleaning</i> .....	30
Tabel 3.6 <i>Tokenizing</i> .....	32
Tabel 3.7 Contoh Dokumen .....	34
Tabel 3.8 Perhitungan TF Pada Setiap Dokumen .....	34
Tabel 3.9 DF Kata Pada Semua Dokumen .....	35
Tabel 3.10 Perhitungan Nilai TF-IDF.....	36
Tabel 3.11 Pengelompokan Nilai TF-IDF Kedalam Label.....	36
Tabel 3.12 Perhitungan Probabilitas NBC .....	37
Tabel 3.13 Contoh Data Training KNN.....	39
Tabel 3.14 Contoh Data Uji KNN .....	39
Tabel 3.15 Perhitungan Jarak <i>Euclidean Distance</i> .....	39
Tabel 3.16 Contoh Data Set Akun <i>Twitter Personal</i> .....	40
Tabel 3.17 Contoh Data Set Kata Validasi .....	41
Tabel 3.18 Validasi Data Set .....	42
Tabel 3.19 Hasil Klasifikasi Akun <i>Twitter Naïve Bayes Classifier</i> .....	42
Tabel 3.20 Hasil Klasifikasi Akun <i>Twitter K – Nearest Neighbor</i> .....	43
Tabel 3.21 Contoh <i>Confusion Matrix Naïve Bayes Classifier</i> .....	44
Tabel 3.22 Contoh <i>Confusion Matrix K – Nearest Neighbor</i> .....	44
Tabel 3.23 Contoh <i>Precision dan Recall Naïve Bayes Classifier</i> .....	46
Tabel 3.24 Contoh <i>Precision dan Recall K – Nearest Neighbor</i> .....	46
Tabel 4.1 Contoh Data Set Akun <i>Twitter Personal</i> .....	58
Tabel 4.2 Contoh Data Set Kata Validasi .....	59
Tabel 4.3 Validasi Data Set .....	60

Tabel 4.4 Hasil Klasifikasi Akun <i>Twitter Naïve Bayes Classifier</i> .....	67
Tabel 4.5 Hasil Klasifikasi Akun <i>Twitter K – Nearest Neighbor</i> .....	68
Tabel 4.6 <i>Confusion Matrix Naïve Bayes Classifier</i> .....	69
Tabel 4.7 <i>Confusion Matrix K – Nearest Neighbor</i> .....	69
Tabel 4.8 <i>Precision dan Recall Naïve Bayes Classifier</i> .....	72
Tabel 4.9 <i>Precision dan Recall K – Nearest Neighbor</i> .....	72
Tabel 4.10 Hasil Klasifikasi Akun <i>Twitter Naïve Bayes Classifier</i> .....	75
Tabel 4.11 Hasil Klasifikasi Akun <i>Twitter K – Nearest Neighbor</i> .....	76
Tabel 4.12 <i>Confusion Matrix Naïve Bayes Classifier</i> .....	76
Tabel 4.13 <i>Confusion Matrix K – Nearest Neighbor</i> .....	77
Tabel 4.14 <i>Precision dan Recall Naïve Bayes Classifier</i> .....	78
Tabel 4.15 <i>Precision dan Recall K – Nearest Neighbor</i> .....	78
Tabel 4.16 Tabel <i>Black Box Testing</i> .....	88



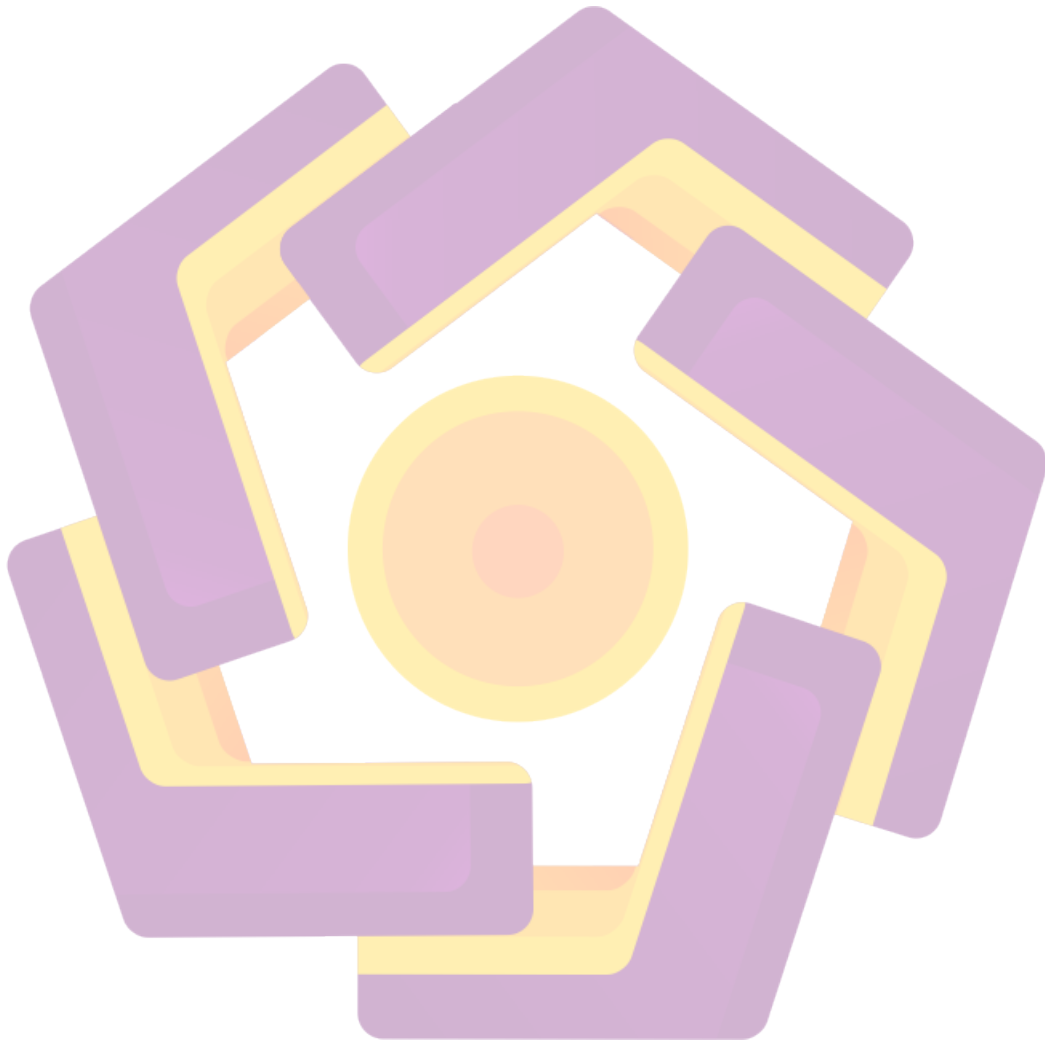


## DAFTAR GAMBAR

Gambar 3.1 Alur Proses Pada Sistem .....	26
Gambar 3.2 Proses <i>Preprocessing</i> .....	28
Gambar 3.3 Proses TF-IDF.....	33
Gambar 3.4 Persamaan Akurasi.....	45
Gambar 3.5 Akurasi <i>Naïve Bayes Classifier</i> .....	45
Gambar 3.6 Akurasi <i>K – Nearest Neighbor</i> .....	45
Gambar 3.7 Dataset Akun, <i>Tweet</i> , dan Label .....	47
Gambar 3.8 Rancangan Antarmuka Halaman <i>Scraping</i> .....	48
Gambar 3.9 Rancangan Antarmuka Halaman Detail Akun.....	49
Gambar 3.10 Rancangan Antarmuka Halaman <i>Preprocessing</i> .....	50
Gambar 3.11 Rancangan Antarmuka Halaman Detail <i>Preprocessing</i> .....	51
Gambar 3.12 Rancangan Antarmuka Halaman Klasifikasi .....	52
Gambar 3.13 Rancangan Antarmuka Halaman Detail Klasifikasi .....	53
Gambar 4.1 Contoh Hasil Evaluasi <i>Confusion Matrix</i> Dari Sistem .....	70
Gambar 4.2 Akurasi <i>Naïve Bayes Classifier</i> .....	71
Gambar 4.3 Akurasi <i>K – Nearest Neighbor</i> .....	71
Gambar 4.4 Diagram Perbandingan Tingkat Akurasi Algoritma .....	73
Gambar 4.5 Diagram Perbandingan Nilai <i>Precision</i> Algoritma.....	73
Gambar 4.6 Diagram Perbandingan Nilai <i>Recall</i> Algoritma .....	74
Gambar 4.7 Akurasi <i>Naïve Bayes Classifier</i> .....	77
Gambar 4.8 Akurasi <i>K – Nearest Neighbor</i> .....	78
Gambar 4.9 Diagram Perbandingan Tingkat Akurasi Algoritma .....	79
Gambar 4.10 Diagram Perbandingan Nilai <i>Precision</i> Algoritma .....	80
Gambar 4.11 Diagram Perbandingan Nilai <i>Recall</i> Algoritma .....	80
Gambar 4.12 Antarmuka Halaman Akun Twitter.....	81
Gambar 4.13 Antarmuka Halaman Detail Akun <i>Twitter</i> .....	82
Gambar 4.14 Antarmuka Halaman <i>Preprocessing</i> .....	83
Gambar 4.15 Antarmuka Halaman Detail <i>Preprocessing</i> .....	84
Gambar 4.16 Antarmuka Halaman Klasifikasi .....	85

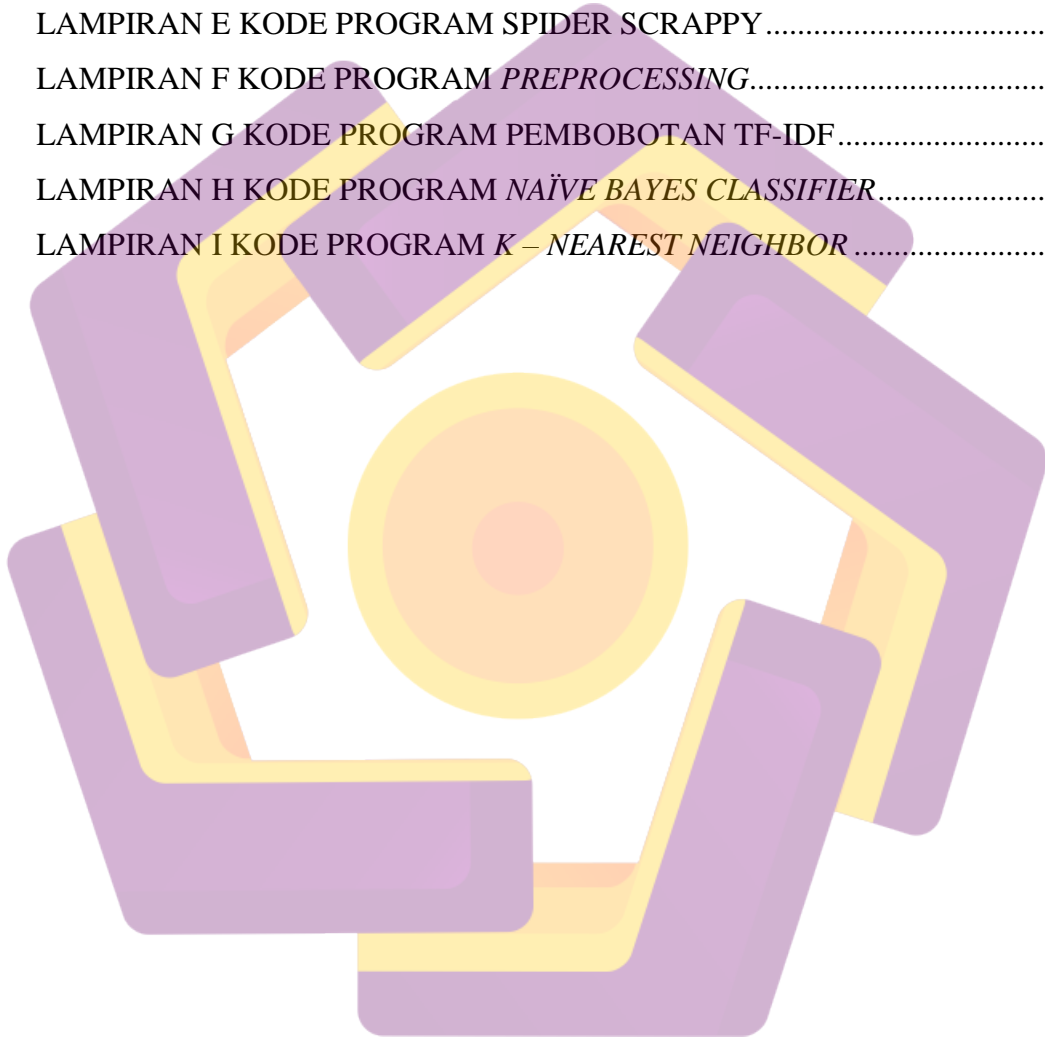
Gambar 4.17 Antarmuka Halaman Detail Klasifikasi .....86

Gambar 4.18 Antarmuka Halaman Evaluasi.....87



## DAFTAR LAMPIRAN

LAMPIRAN A CONTOH DATA <i>TWEET</i> .....	1
LAMPIRAN B CONTOH DATA AKUN.....	5
LAMPIRAN C CONTOH DATA TRAINING KNN .....	7
LAMPIRAN D CONTOH DATA KATA.....	8
LAMPIRAN E KODE PROGRAM SPIDER SCRAPPY.....	11
LAMPIRAN F KODE PROGRAM <i>PREPROCESSING</i> .....	15
LAMPIRAN G KODE PROGRAM PEMBOBOTAN TF-IDF.....	18
LAMPIRAN H KODE PROGRAM <i>NAÏVE BAYES CLASSIFIER</i> .....	20
LAMPIRAN I KODE PROGRAM <i>K – NEAREST NEIGHBOR</i> .....	22



## INTISARI

*Twitter* merupakan salah satu media sosial yang umum digunakan oleh masyarakat terutama di Indonesia. Angka ini menunjukkan bahwa banyak masyarakat Indonesia yang menggunakan media sosial *Twitter* untuk berkomunikasi ataupun menyampaikan pendapatnya. Komunikasi yang disampaikan melalui *Tweet* ini dapat diolah untuk mendapatkan informasi baru seperti karakteristik seseorang menggunakan pendekatan DISC.

Dengan mengetahui karakter individu tersebut, hal ini bisa digunakan sebagai tolak ukur seperti saat melakukan seleksi karyawan. Sehingga pihak HRD (*Human Resource Department*) tidak perlu mengeluarkan tenaga lebih untuk mengetahui karakter individu tersebut di dunia maya.

Namun untuk melakukan hal ini, diperlukan sebuah metode lain yaitu Data Mining. Untuk dapat mengklasifikasikan tiap *tweet* dari individu tersebut diperlukan algoritma yang terdapat dalam *Data Mining* seperti *K – Nearest Neighbor* dan *Naïve Bayes Classifier*.

Pada Penelitian ini dengan menggunakan 120 akun *Twitter* yang sudah divalidasi oleh psikolog secara manual akan membandingkan tingkat akurasi algoritma *K – Nearest Neighbor* dan *Naïve Bayes Classifier*, didapati hasil bahwa algoritma *K – Nearest Neighbor* memiliki tingkat akurasi 28.33% dan *Naive Bayes Classifier* 34.16%, sehingga dapat disimpulkan bahwa algoritma *Naive Bayes Classifier* lebih unggul dibanding *K – Nearest Neighbor* sebesar 5.83% dalam hal tingkat akurasi.

**Kata kunci :** Analisis karakter, Twitter, teori DISC, Algoritma, *Naive Bayes Classifier*, *K – Nearest Neighbor*, Klasifikasi, Perbandingan.

## **ABSTRACT**

*Twitter is a social media that is commonly used by the public, especially in Indonesia. This figure shows that many Indonesian people use Twitter social media to communicate or express their opinions. Communication delivered through this Tweet can be processed to obtain new information such as the characteristics of someone using the DISC approach.*

*By knowing the individual character, this can be used as a benchmark like when selecting employees. So that the HRD (Human Resource Department) does not need to spend more energy to find out the individual character in cyberspace.*

*But to do this, another method is needed, namely Data Mining. To be able to classify each tweet from the individual requires an algorithm contained in Data Mining such as K - Nearest Neighbor and Naïve Bayes Classifier.*

*In this study using 120 Twitter accounts that have been validated by psychologists manually comparing the accuracy of the K-Nearest Neighbor and Naïve Bayes Classifier algorithm, the results show that the K-Nearest Neighbor algorithm has an accuracy rate of 28.33% and Naive Bayes Classifier 34.16%, so it can be concluded that the Naïve Bayes Classifier algorithm is superior to the K-Nearest Neighbor by 5.83% in terms of accuracy.*

**Keyword** : *Character Analysis, Twitter, DISC theory, Algorithm, Naive Bayes Classifier, K – Nearest Neighbor, classification, Comparison.*