

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST  
NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI  
KARAKTER INDIVIDU PADA AKUN TWITTER**

**SKRIPSI**



disusun oleh

**Aristin Chusnul Khotimah**

**17.11.1535**

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

**2022**

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST  
NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI  
KARAKTER INDIVIDU PADA AKUN TWITTER**

**SKRIPSI**

untuk memenuhi sebagian persyaratan mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Aristin Chusnul Khotimah**

**17.11.1535**

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

**2022**

**PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI KARAKTER INDIVIDU PADA AKUN TWITTER**

yang dipersiapkan dan disusun oleh

**Aristin Chusnul Khotimah**

**17.11.1535**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 Januari 2022

**Dosen Pembimbing.**

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

**NIK. 1903020037**

**PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI KARAKTER INDIVIDU PADA AKUN TWITTER**

yang dipersiapkan dan disusun oleh

**Aristin Chusnul Khotimah**

**17.11.1535**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 23 Februari 2022

Susun Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Anggit Dwi Hartanto, M.Kom**

**NIK. 190302163**

**Kusnawi, S.Kom, M. Eng.**

**NIK. 190302112**

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

**NIK. 190302037**

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer

Tanggal 23 Februari 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

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

**NIK. 190302096**

## PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan di dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu instansi 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.

Seperti sesuatu yang tercantum dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 28 Maret 2022



Azzatin Chusmi Khotimah

NIM. 17.11.1535

## MOTTO

*"...Tersenyumlah, senyummu bisa jadi obat bagi luka yang tak kau sangka-sangka."*

(Seni Melipat Luka)

*"Berjuangmu itu tidak akan sia-sia didepan Tuhan."* (Gayatri H. Putri)

*"Mamusia itu tercipta dengan kapasitas yang terbatas namun dengan emosi yang luas. Tak perlu selalu berlari, terkadang kamu juga perlu berhenti."*

(Gayatri H. Putri)

*"Jangan mau kalah sama suara-suara di kepalamu yang bilang kamu nggak nggak akan pernah bisa. Kamu lebih besar dari itu. Kamu adalah semesta untuk dirimu sendiri."*

(Diego Christian)



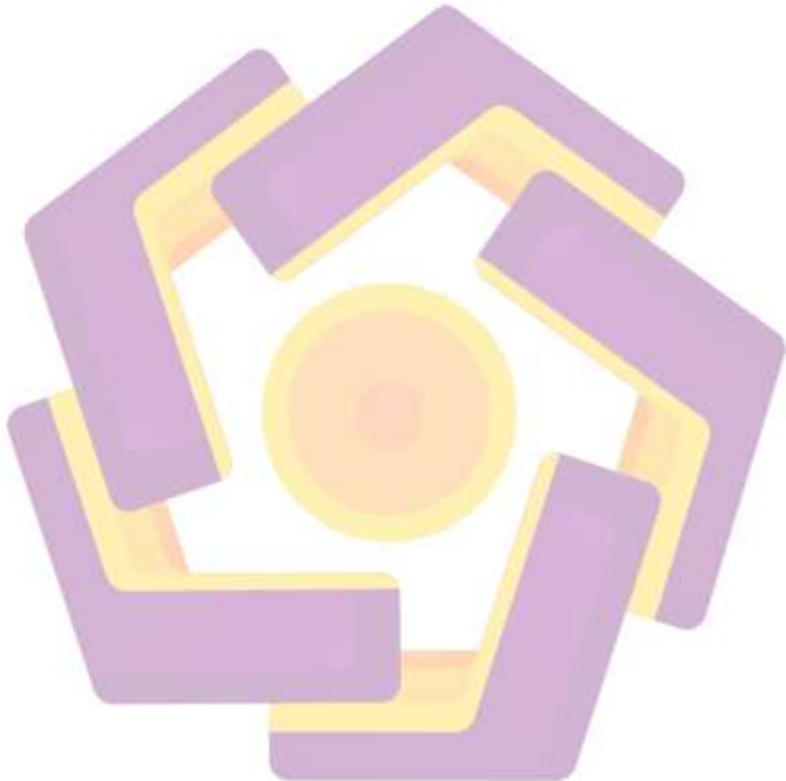
## PERSEMBAHAN

Puji syukur peneliti ucapkan kepada Allah SWT atas kelancaran untuk menyelesaikan skripsi ini dengan baik. Dengan ini peneliti mengucapkan terima kasih kepada semua pihak yang telah membantu dalam menyelesaikan skripsi ini dan dengan ini saya persembahkan skripsi kepada:

1. Kedua orang tua tercinta, almh mama dan papa yang selalu berusaha memberikan segala fasilitas yang sangat lengkap dari saya lahir didunia ini. Semoga dengan selesainya skripsi ini almh mama dan papa bangga dan bahagia melihat anak satu-satunya bisa ngelewatin ini semua.
2. Dosen pembimbing, Ibu Prof. Dr. Ema Utami, S.Si., M.Kom yang telah membimbing saya untuk menyelesaikan skripsi.
3. Sahabat seperjuangan, Nurul Khairani Moha, Nilam Pangestika, Amanda Wahyu Puspita Sari, Fharah Fakhira Maharani, Ninda Putri Weliza dan Kesya C. Tehupeiori. Terimakasih sudah selalu saling mendengar, mensupport dan mendoakan kebaikan dalam keadaan apapun.
4. Teman sekelas, Faradila Mala Putri. Terimakasih sudah mau berteman dengan tulus, sudah mau saling mendengar, mensupport dan mendoakan kebaikan dalam keadaan apapun.
5. Sahabat seperjuangan, Dinda Zuliani Madjid dan Halimah Aprilia Sudirman. Terimakasih sudah selalu berkunjung ke Yogyakarta dan selalu mensupport kebaikan dalam keadaan apapun.



6. Teman seperjuangan, Putra Mukhamad Mei Bimantara, Ahmad Fadhil Ridwan, Muh. Fuad Fauzi, Ahmad Mirza W, Fikri Abdilah dan Abdul Azhim A. Terimakasih sudah menjadi teman sedari SMA, selalu ngelawak untuk menjeda beban pikiran dan jadi tempat bertukar pikiran.
7. Semua pihak yang tidak bisa disebutkan satu-persatu dalam membantu skripsi ini.





## KATA PENGANTAR

Segala puji dan syukur senantiasa penulis panjatkan kepada Allah SWT, atas karunia dan rahmat-Nya penulis dapat menyelesaikan skripsi berjudul “PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI KARAKTER INDIVIDU PADA AKUN TWITTER” sebagai syarat kelulusan untuk mendapat gelar Sarjana Komputer di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terimakasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega PD, M.Kom selaku Ketua Program Studi Informatika.
4. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom selaku Dosen Pembimbing.
5. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu pada bangku perkuliahan.
6. Orang tua yang selalu membantu dengan doa kepada penulis selama menempuh kuliah dan penyusunan skripsi.
7. Mentor yang selalu memberikan saran dan masukan untuk menyelesaikan skripsi ini.

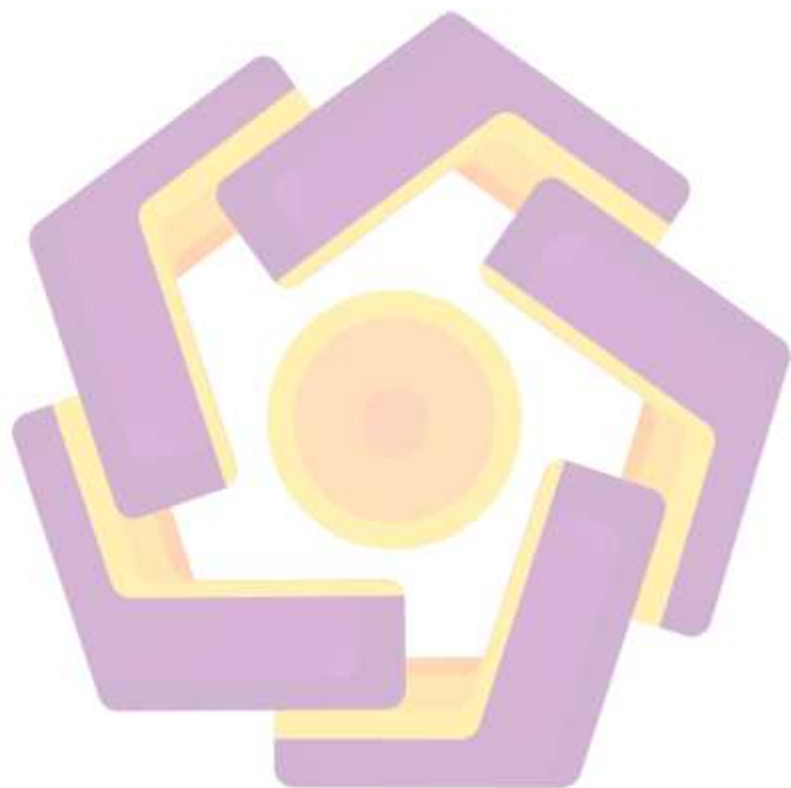
Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh sebab itu penulis mengharapkan kritik dan saran yang bersifat membangun dari semua pihak demi kesempurnaan skripsi ini. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi kemajuan ilmu pengetahuan dan tentunya masyarakat.

## DAFTAR ISI

PERSETUJUAN.....	I
PENGESAHAN.....	II
PERNYATAAN.....	<b>Error! Bookmark not defined.</b>
MOTTO.....	IV
PERSEMBAHAN.....	V
KATA PENGANTAR.....	VII
DAFTAR ISI.....	VIII
DAFTAR LABEL.....	XI
DAFTAR GAMBAR.....	XIV
DAFTAR LAMPIRAN.....	XVI
INTISARI.....	XVII
ABSTRACT.....	XIX
BAB I PENDAHULUAN.....	1
1.1    Latar Belakang.....	1
1.2    Rumusan Masalah.....	3
1.3    Batasan Masalah.....	4
1.4    Tujuan Penelitian.....	4
1.5    Manfaat Penelitian.....	5
1.6    Metodologi Penelitian.....	5
1.7    Sistematika Penulis.....	6
BAB II LANDASAN TEORI.....	8
2.1    Tinjauan Pustaka.....	8
2.2    Dasar Teori.....	11
2.2.1    Sosial Media.....	11
2.2.1.1    Twitter.....	12
2.2.2    Karakter.....	13
2.2.3    Teori DISC (Dominance, Influence, Steadiness, Compliance).....	13
2.2.4    Knowledge Discovery in Database (KDD).....	15
2.2.5    Data Mining.....	16
2.2.6    Crawling.....	20
2.2.7    Preprocessing.....	21
2.2.7.1    Case Folding.....	21
2.2.7.2    Cleansing.....	22

2.2.7.3	Stemming .....	22
2.2.7.4	Stopword Removal .....	22
2.2.7.5	Tokenizing.....	22
2.2.8	Pembobotan TF-IDF .....	23
2.2.8.1	Trem Frequency (TF) .....	23
2.2.8.2	Document Frequency (DF).....	23
2.2.8.3	Invers Document Frequency (IDF) .....	23
2.2.8.4	Trem Frequency – Invers Document Frequency (TF-IDF).....	24
2.2.9	Algoritma Naïve Bayes classifier.....	24
2.2.10	Algoritma K-Nearest Neighbor .....	26
2.2.11	Euclidean Distance .....	27
2.2.12	Support Vector Machine .....	28
2.2.13	Confusion Matrix .....	32
<b>BAB III</b>	<b>ANALISIS DAN PERANCANGAN .....</b>	<b>34</b>
3.1	Sistem.....	34
3.2	Gambaran Umum Sistem .....	35
3.3	Analisis Algoritma .....	36
3.4	Analisis Kebutuhan Data.....	72
3.5	Perancangan Antarmuka .....	73
<b>BAB IV</b>	<b>IMPLEMENTASI DAN PEMBAHASAN .....</b>	<b>78</b>
4.1	Implementasi .....	78
4.2	Implementasi Pengumpulan Data Tweet.....	78
4.3	Implementasi Pelabelan Dataset.....	79
4.4	Implementasi Preprocessing.....	80
4.5	Implementasi Pembobotan dan Klasifikasi .....	82
4.6	Evaluasi Confusion Matrix.....	85
4.7	Evaluasi .....	87
4.8	Implementasi Antarmuka .....	93
4.9	Pengujian Sistem.....	99
<b>BAB V</b>	<b>PENUTUP .....</b>	<b>100</b>
5.1	Kesimpulan.....	100
5.2	Saran.....	100
<b>DAFTAR PUSTAKA .....</b>	<b>102</b>	
<b>LAMPIRAN A.....</b>	<b>1</b>	
<b>LAMPIRAN B .....</b>	<b>4</b>	

LAMPIRAN C.....	7
LAMPIRAN D.....	10
LAMPIRAN E.....	13
LAMPIRAN F.....	16



## DAFTAR LABEL

Tabel 2.1 : Perbandingan Pustaka.....	11
Tabel 2.2 : Kernel yang umum dipake dalam SVM .....	32
Tabel 2.3 : Confusion Matrix .....	32
Tabel 3.1 : Data Set Tweet.....	36
Tabel 3.2 : Pelabelan Trem .....	37
Tabel 3.3 : Pelabelan Akun.....	38
Tabel 3.4 : Proses Case Folding.....	39
Tabel 3.5 : Proses Cleaning .....	39
Tabel 3.6 : Proses Stemming .....	40
Tabel 3.7 : Proses Stopword .....	41
Tabel 3.8 : Proses Tokenizing.....	41
Tabel 3.9 : Proses Trem Dokumen .....	42
Tabel 3.10 : Proses Term Frequency .....	43
Tabel 3.11 : Proses Inverse Document Frequency.....	44
Tabel 3.12 : Proses TF-IDF.....	45
Tabel 3.13 : Proses Pelabelan Trem.....	45
Tabel 3.14 : Proses Naïve Bayes Classifier .....	46
Tabel 3.15 : Hasil Pelabelan Data Training .....	47
Tabel 3.16 : Proses K-Nearest Neighbor .....	47
Tabel 3.17 : Hasil Pelabelan Data Training .....	49



Tabel 3.18 : Hasil $x_i - x_i$ .....	51
Tabel 3.19 : Hasil Perhitungan Panjang Vektor.....	51
Tabel 3.20 : Hasil Perhitungan $y_i$ .....	53
Tabel 3.21 : Hasil Perhitungan $y_i y_i$ .....	53
Tabel 3.22 : Hasil Perhitungan $E_i$ .....	54
Tabel 3.23 : Hasil Perhitungan $\delta a_i$ .....	55
Tabel 3.24 : Hasil Perhitungan $\alpha_i$ .....	55
Tabel 3.25 : Hasil Perhitungan Dot Product.....	57
Tabel 3.26 : Hasil Pelabelan Akun .....	59
Tabel 3.27: Hasil Pelabelan Trem.....	59
Tabel 3.28 : Hasil Psikolog.....	59
Tabel 3.29 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma Naïve Bayes Classifier.....	60
Tabel 3.30 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma K-Nearest Neighbor .....	61
Tabel 3.31 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma Support Vector Machine.....	61
Tabel 3.32 : Confusion Matrix Naïve Bayes Classifier .....	62
Tabel 3.33 : Hasil Presisi Naïve Bayes Classifier.....	64
Tabel 3.34 : Hasil Recall Naïve Bayes Classifier.....	65
Tabel 3.35 : Confusion Matrix K-Nearest Neighbor .....	65
Tabel 3.36 : Hasil Presisi K-Nearest Neighbor.....	67

Tabel 3.37 : Hasil Recall K-Nearest Neighbor .....	68
Tabel 3.38 : Confusion Matrix Support Vector Machine .....	69
Tabel 3.39 : Hasil Presisi Support Vector Machine .....	71
Tabel 3.40 : Hasil Recall Support Vector Machine .....	72
Tabel 4.1 : Contoh Data Set Akun Twitter .....	79
Tabel 4.2 : Contoh Data Set Trem .....	80
Tabel 4.3 : Data Set Psikolog .....	80
Tabel 4.4 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma Naïve Bayes Classifier .....	86
Tabel 4.5 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma K-Nearest Neighbor .....	86
Tabel 4.6 : Hasil Klasifikasin Akun Twitter Menggunakan Algoritma Support Vector Machine .....	87
Tabel 4.7 : Confusion Matrix Naïve Bayes Classifier .....	87
Tabel 4.8 : Confusion Matrix K-Nearest Neighbor .....	88
Tabel 4.9 : Confusion Matrix Support Vector Machine .....	88
Tabel 4.10 : Presisi dan Recall Naïve Bayes Classifier .....	91
Tabel 4.11 : Presisi dan Recall K-Nearest Neighbor .....	91
Tabel 4.12 : Presisi dan Recall Support Vector Machine .....	91
Tabel 4.13 : Pengujian black box .....	99



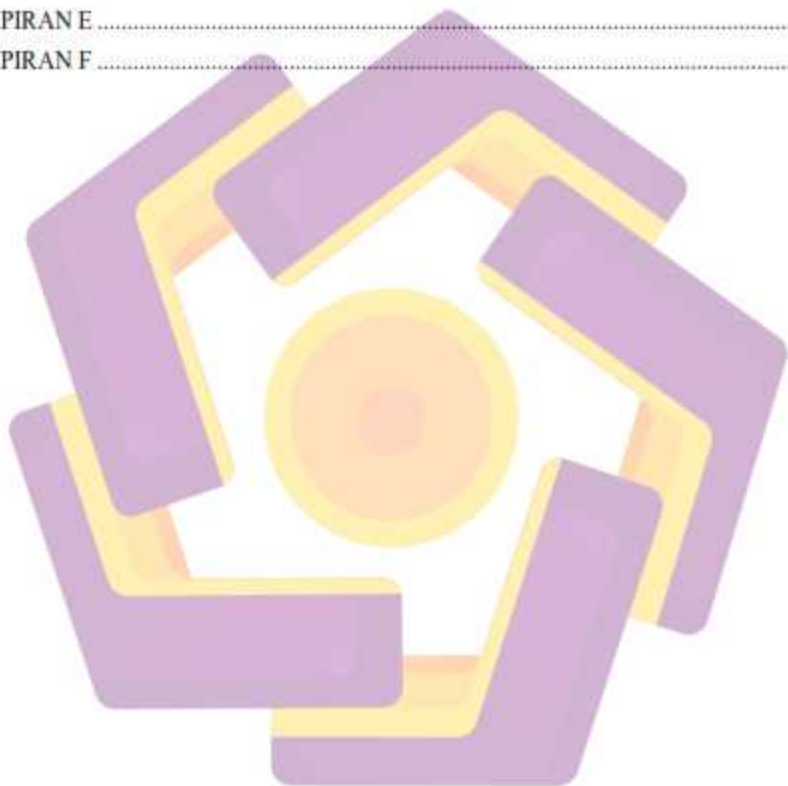
## DAFTAR GAMBAR

Gambar 3.1: Flowchart Pada Sistem.....	36
Gambar 3.2 : Flowchart Pada Preprocessing .....	38
Gambar 3.3 : Flowchart Pada TF-IDF .....	42
Gambar 3.4 : Data Sebelum Preprocessing.....	73
Gambar 3.5 : Antarmuka Halaman Beranda.....	74
Gambar 3.6 : Antarmuka Halaman Search Akun .....	74
Gambar 3.7 : Antarmuka Halaman Detail Search Akun.....	75
Gambar 3.8 : Antarmuka Halaman Preprocessing Data .....	75
Gambar 3.9 : Antarmuka Halaman Detail Preprocessing Data .....	76
Gambar 3.10: Antarmuka Halaman Klasifikasi.....	76
Gambar 3.11 : Antarmuka Halaman Detail Klasifikasi.....	77
Gambar 4.1 : Code Program Crawling Data Twitter .....	78
Gambar 4.2 : Code Case Folding.....	80
Gambar 4.3 : Code Program Hapus Tanda Baca .....	81
Gambar 4.4 : Code Program Cleansing .....	81
Gambar 4.5 : Code Program Stemming.....	81
Gambar 4.6 : Code Program Stopword Removal .....	82
Gambar 4.6 : Code Program Tokenzing .....	82
Gambar 4.7 : Code Program Algoritma Naïve Bayes Classifier .....	84
Gambar 4.8 : Code Program Algoritma K-Nearest Neighbor .....	84

Gambar 4.9 : Code Program Algoritma Support Vector Machine .....	85
Gambar 4.10 : Code Program Confusion Matrix.....	88
Gambar 4.11 : Hasil dari Evaluasi Confusion Matrix.....	89
Gambar 4.12 : Akurasi Algoritma Naïve Bayes Classifier.....	90
Gambar 4.13 : Akurasi Algoritma K-Nearest Neighbor.....	90
Gambar 4.14 : Akurasi Algoritma Support Vector Machine .....	90
Gambar 4.15 : Diagram Pembeding Tingkat Akurasi Algoritma .....	92
Gambar 4.16 : Diagram Pembeding Nilai Presisi Algoritma .....	92
Gambar 4.17 : Diagram Pembeding Nilai Recall Algoritma.....	92
Gambar 4.18 : Antarmuka Halaman Beranda.....	93
Gambar 4.19 : Antarmuka Halaman Search .....	94
Gambar 4.20 : Antarmuka Halaman Detail Search .....	95
Gambar 4.21 : Antarmuka Halaman Preprocessing Data .....	95
Gambar 4.21 : Antarmuka Halaman Detail Preprocessing Data .....	96
Gambar 4.22 : Antarmuka Halaman Klasifikasi.....	96
Gambar 4.23 : Antarmuka Halaman Detail Klasifikasi .....	97
Gambar 4.24 : Antarmuka Halaman Confusion Matrix.....	98
Gambar 4.25 : Antarmuka Halaman Grafik.....	98

## DAFTAR LAMPIRAN

LAMPIRAN A.....	1
LAMPIRAN B.....	4
LAMPIRAN C.....	7
LAMPIRAN D.....	10
LAMPIRAN E.....	13
LAMPIRAN F.....	16



## INTISARI

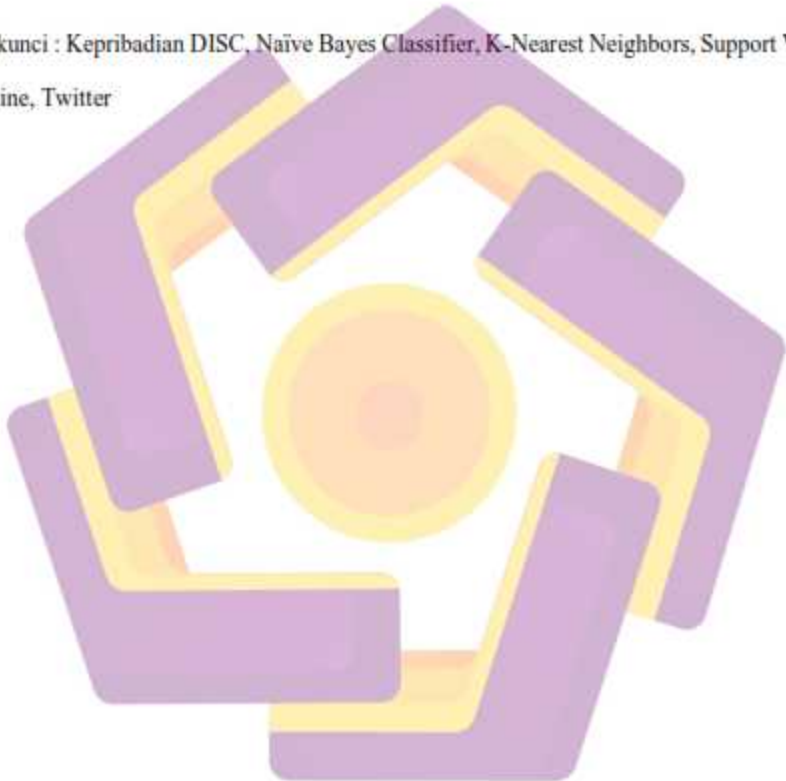
Pada era digital saat ini, masyarakat dapat memanfaatkan kemudahan dan keefektifan dalam berinteraksi antara satu sama lain. Kegiatan online yang populer di Indonesia adalah penggunaan media sosial. Dengan adanya jaringan internet dan perkembangan media sosial tersebut, kini siapapun dapat terhubung dengan siapapun di penjuru dunia manapun. Dengan 120 juta pengguna diantaranya menggunakan ponsel untuk terhubung dengan akun media sosial. Twitter adalah media sosial yang memungkinkan orang untuk membangun komunikasi antar pengguna dan dapatkan informasi terbaru atau berita. Informasi yang didapatkan dari twitter dapat olah untuk mendapatkan karakteristik seseorang menggunakan metode DISC, DISC merupakan model perilaku yang membantu setiap manusia mengapa seseorang melakukan apa yang dia lakukan.

Metode penelitian ini dilakukan pengumpulan data dari data twitter menggunakan crawling dengan library yang disediakan oleh python yaitu tweepy yang berupa seluruh tweet dari 130 akun personal. Setelah data diperoleh maka dilakukan pelabelan data secara manual oleh psikolog kemudian akan mengklasifikasikan akun twitter kedalam metode DISC digunakan ketiga algoritma yaitu algoritma Naïve Bayes Classifier, K-Nearest Neighbor dan Support Vector Machine. Setelah mengetahui karakter per-akun kemudian membandingkan kinerja ketiga algoritma tersebut dengan tahapan preprocessing, pembobotan TF-IDF dan mengklasifikasikan akun twitter kedalam metode DISC dapat diketahui algoritma mana yang lebih tinggi tingkat akurasi.

Hasil penelitian ini membandingkan tingkat akurasi algoritma Naïve Bayes Classifier, K-Nearest Neighbor dan Support Vector Machine. Algoritma Naïve Bayes

Classifier memiliki tingkat akurasi sebesar 31.5%, K-Nearest Neighbor memiliki tingkat akurasi sebesar 23.8%, sedangkan Support Vector Machine memiliki tingkat akurasi sebesar 28.4%. Sehingga kesimpulannya algoritma Naïve Bayes Classifier lebih unggul dibandingkan dengan algoritma K-Nearest Neighbor dan Support Vector Machine, namun Support Vector Machine perbedaan tingkat akurasi sebesar 4.7% dengan algoritma K-Nearest Neighbor.

Kata kunci : Kepribadian DISC, Naïve Bayes Classifier, K-Nearest Neighbors, Support Vector Machine, Twitter





## ABSTRACT

*In current's digital era, people can take advantage of the ease and effectiveness of interacting with each other. The most popular online activity in Indonesia is the use of social media. With the internet network and the development of social media, now anyone can connect with anyone anywhere in the world. With 120 million of them use mobile phones to connect with social media accounts. Twitter is a social media that allows people to build communication between users and get the latest information or news. Information obtained from twitter can be processed to get the characteristics of a person using the DISC method, DISC is a behavioral model that helps every human being why someone does.*

*This research method is collecting data from twitter data using crawling with a library provided by python, namely tweepy in the form of all tweets from 130 personal accounts. After the data is obtained, the psychologist will label the data manually and then classify the twitter account into the DISC method using three algorithms, namely the Naïve Bayes Classifier, K-Nearest Neighbor and Support Vector Machine algorithms. After knowing the characters per account and then comparing the performance of the three algorithms with the preprocessing stage, TF-IDF weighting and classifying twitter accounts into the DISC method, it can be seen which algorithm has a higher accuracy level.*

*The results of this research is compare the accuracy of the Naïve Bayes Classifier, K-Nearest Neighbor and Support Vector Machine. The Naïve Bayes Classifier algorithm has an accuracy rate of 31.5%, K-Nearest Neighbor has an accuracy rate of 23.8%, while the Support Vector Machine has an accuracy rate of 28.4%. So in conclusion, the Naïve Bayes Classifier algorithm is superior to the K-Nearest Neighbor algorithm and Support Vector Machine, but the Support Vector Machine has a 4.7% difference in accuracy with the K-Nearest Neighbor algorithm.*

**Keyword** : *Personality DISC, Naïve Bayes Classifier, K-Nearest Neighbors, Support Vector Machine, Twitter*

