

**IMPLEMENTASI METODE RANDOM FOREST CLASSIFIER UNTUK
PREDIKSI KEPRIBADIAN DENGAN PENDEKATAN DISC
BERDASARKAN CAPTION INSTAGRAM BAHASA INDONESIA**

SKRIPSI



disusun oleh

Ahmad Amdad Bariro

18.12.0780

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

**IMPLEMENTASI METODE RANDOM FOREST CLASSIFIER UNTUK
PREDIKSI KEPERIBADIAN DENGAN PENDEKATAN DISC
BERDASARKAN CAPTION INSTAGRAM BAHASA INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Ahmad Amdad bariro

18.12.0780

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE RANDOM FOREST CLASSIFIER
UNTUK PREDIKSI KEPRIBADIAN DENGAN PENDEKATAN DISC
BERDASARKAN CAPTION INSTAGRAM BAHASA INDONESIA**

Yang dipersiapkan dan disusun oleh :

Ahmad Amdad Bariro
18.12.0780

Telah disetujui oleh Dosen Pembimbing Skripsi
Pada tanggal 22 November 2021

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

PENGESAHAN
SKRIPSI
IMPLEMENTASI METODE RANDOM FOREST CLASSIFIER
UNTUK PREDIKSI KEPRIBADIAN DENGAN PENDEKATAN DISC
BERDASARKAN CAPTION INSTAGRAM BAHASA INDONESIA

Yang dipersiapkan dan disusun oleh

Ahmad Amdad Bariro
18.12.0780

Telah dipertahankan didepan Dewan Penguji
Pada tanggal 22 November 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

Ika Nur Fajri, M.Kom.

NIK. 190302268

Robert Marco, M.T.

NIK. 190302228

Skripsi ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar Sarjana Komputer
Tanggal 28 November 2021

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 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, 22 November 2021



Ahmad Amdad Bariro

NIM. 18.12.0780

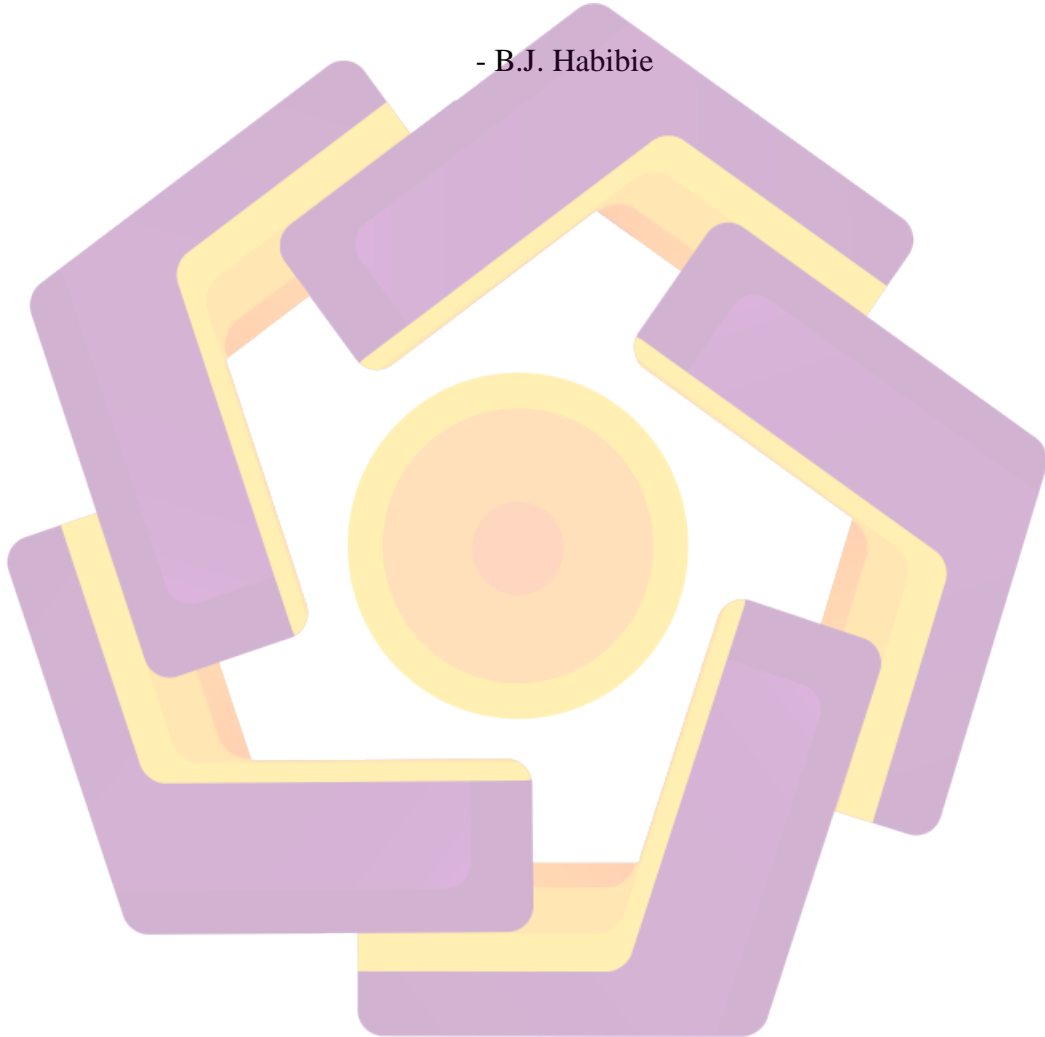
MOTTO

(Man Jadda Wajada)

“Barang siapa yang bersungguh-sungguh, maka ia akan berhasil”

“Keberhasilan bukanlah milik orang yang pintar, keberhasilan adalah kepunyaan mereka yang senantiasa berusaha.”

- B.J. Habibie

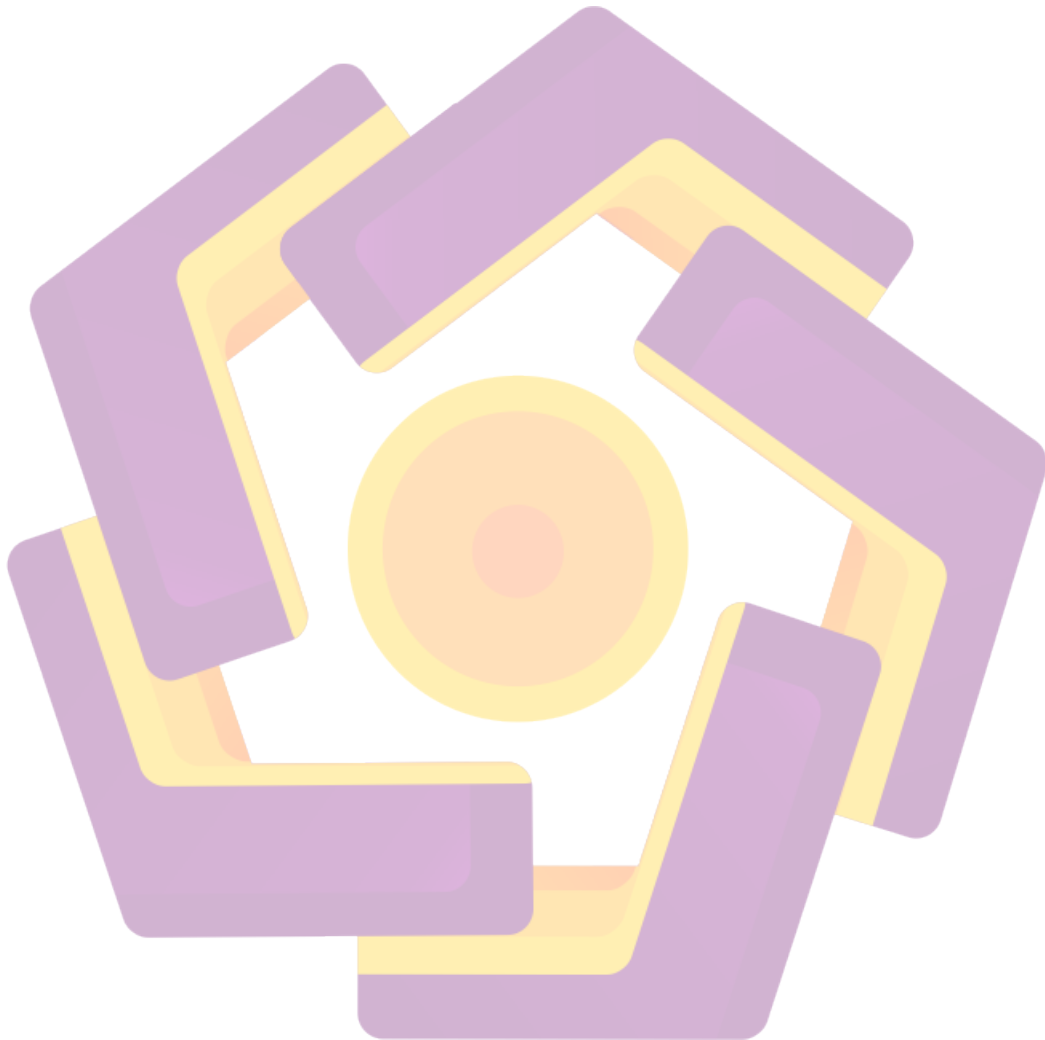


PERSEMBAHAN

Alhamdulillah, segala puji bagi Allah SWT dengan segala kuasa dan rahmat-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan lancar dan sesuai dengan target serta mendapatkan hasil yang terbaik. Terima kasih juga tak lupa saya sampaikan kepada orang-orang yang telah memberikan semangat secara moril dan membantu dalam segala hal. Skripsi ini saya persembahkan kepada :

1. Allah SWT yang dengan segala rahmat dan karunia-Nya lah seluruh rangkaian proses pengerjaan skripsi ini berjalan dengan lancar, dan dapat selesai tepat waktu serta memperoleh hasil yang maksimal.
2. Kedua orang tua saya yang senantiasa mendoakan, memberikan semangat, dan selalu mendukung dalam segala bidang yang saya tekuni.
3. Keluarga besar Bani Samsir dan Mamiq Hawari, yang selalu memberikan dukungan, semangat, serta nasihat-nasihat.
4. Bapak ibu dosen yang telah memberikan banyak ilmu dan pengalaman selama perkuliahan.
5. Bapak Anggit Dwi Hartato, M.Kom. yang telah membimbing selama penelitian baik dalam perlombaan dan skripsi.
6. Bagas Wahyu Ikhsandi, sahabat dan rekan seperjuangan sejak awal perkuliahan, yang telah berbagi banyak hal baik itu ilmu, pengalaman, masukan, dan dukungan baik di dalam dan di luar perkuliahan.
7. Rekan - rekan kelas 18 Sistem Informasi 04, rekan seperjuangan yang telah berbagi ilmu, belajar, dan bercanda tawa bersama.
8. Beasiswa Unggulan Kemendikbud yang telah membantu saya secara finansial.
9. Mas Bayu Permana Sejati, selaku mentor sekaligus role model, yang telah membimbing, memberikan banyak kritik, saran, dan masukan.
10. Keluarga Besar Forum Asisten, rekan-rekan seperjuangan, Arbi, Bagas, Renaldi, Yustika, Beta, Mitha, Anggita, Dwi, Lia, Dini, Iza, Handa, Nenden.
11. Teman – teman grup All in One, Ali, Bagas, Febri, Wulan, Irma, yang selalu membantu, dan memberi masukan selama perkuliahan.

12. Teman – teman kontrakan Tongkol, Bayu, Hasyim, Rama, dll, yang selalu membantu, main bersama, dan berbagi canda tawa.
13. Seluruh pihak yang tidak dapat saya sebutkan satu per satu, terimakasih atas segala bantuan, doa, dan dukungannya sehingga terselesaikan skripsi ini.



KATA PENGANTAR

Puji Syukur kita panjatkan kepada Allah atas berkat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini tepat pada waktunya dengan judul **“Implementasi Metode Random Forest Classifier Untuk Prediksi Kepribadian Dengan Pendekatan Disc Berdasarkan Caption Instagram Bahasa Indonesia”**.

Skripsi ini disusun untuk memenuhi salah satu persyaratan dalam menyelesaikan Program Strata-I Sistem Informasi di Universitas Amikom Yogyakarta. Selama mengikuti pendidikan Srtata-I Sistem Informasi sampai dengan proses penyelesaian skripsi, berbagai pihak telah memberikan fasilitas, membantu, membina, dan membimbing peulis untuk itu khususnya kepada:

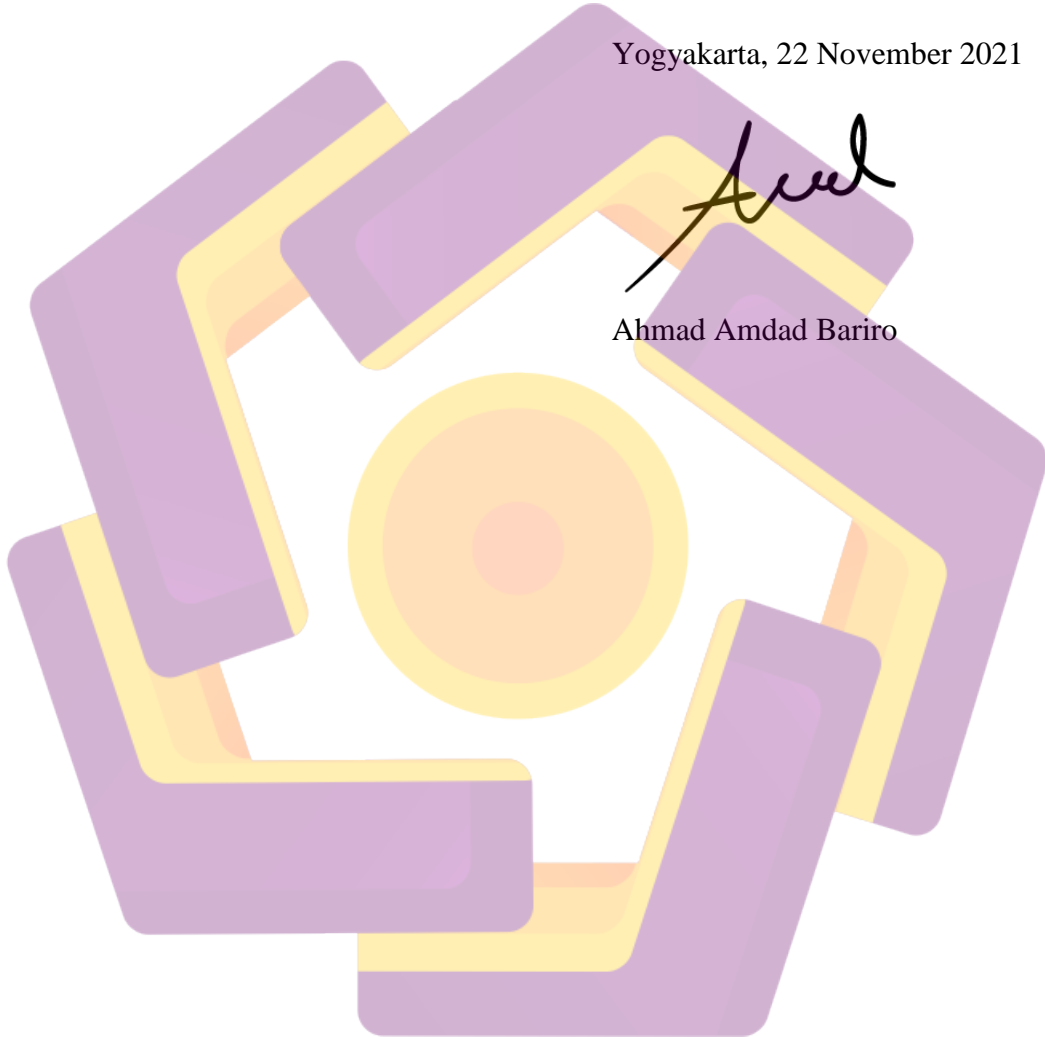
1. Bapak Prof. Dr. M. Suyanto , MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M.Kom. Selaku Kepala Program Studi Sistem Informasi Universitas Amikom Yogyakarta, sekaligus ddosen pembimbing saya yang selalu bijaksana memberikan bimbingan nasehat serta waktu nya selama penulisan skripsi ini.
4. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah banyak memberikan ilmunya selama kuliah.
5. Keluarga Besar Pengurus Forum asisten Universitas Amikom Yogyakarta yang telah memberikan motivasi dan dorongan kepada saya.
6. Teman - teman 18 Sistem Informasi 04 yang sudah berjuang bersama – sama, membagi ilmu, canda tawa, dan kebersamaannya selama perkuliahan.
7. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Tuhan memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi sayadan kita semua.

Yogyakarta, 22 November 2021



Ahmad Amdad Bariro

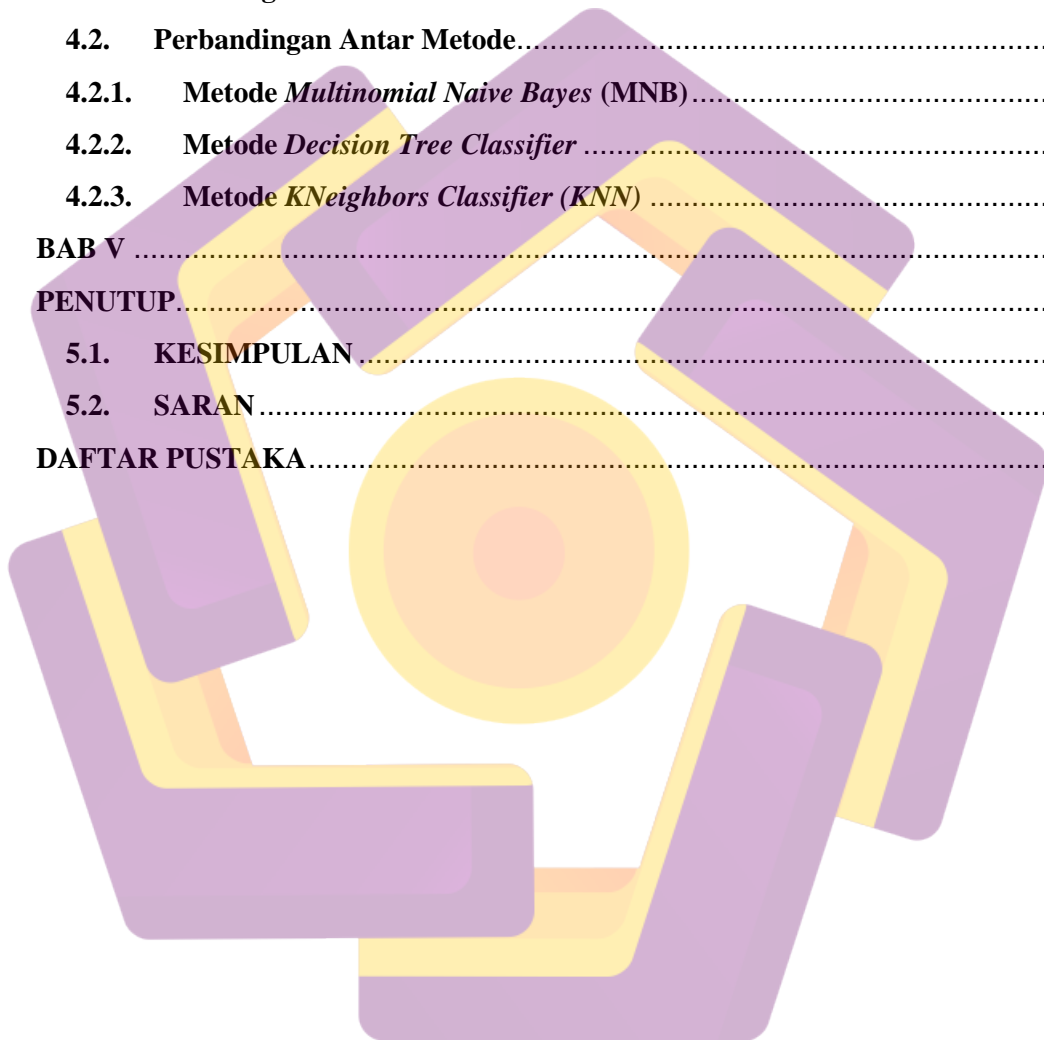


DAFTAR ISI

JUDUL	1
PERSETUJUAN	ii
PERSEMBAHAN	vi
KATA PENGANTAR	viii
INTISARI	xv
ABSTRACT	xvi
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I	xv
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Maksud Dan Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
1.6. Metode Penelitian	4
1.6.1. Metode Studi Literatur	4
1.6.2. Metode Pengumpulan Data	5
1.6.3. Metode Analisis	5
1.6.4. Implementasi	5
1.6.5. Testing	5
1.7. Sistematika Penulisan	5
BAB II	7
LANDASAN TEORI	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori	10
2.2.1. Text Mining	10
2.2.2. NLP (Natural Language Processing)	11
2.2.3. Instagram	11
2.2.3.1. Instagram Scraper	12

2.2.4.	Random Forest Classifier	12
2.2.5.	Pembersihan Data	13
2.2.5.1.	Cleansing	13
2.2.5.2.	Case Folding	14
2.2.5.3.	Stopword Removal	14
2.2.5.4.	Tokenizing	14
2.2.6.	Term Frequency Inverse Document Frequency (TF-IDF)	14
2.2.7.	Testing	15
2.2.7.1.	Confusion Matrix	15
2.2.8.	Word Clouds	16
BAB III		18
METODOLOGI PENELITIAN		18
3.1.	Alur Penelitian	18
3.2.	Sumber Data dan Variabel Penelitian	18
3.2.1.	Data Set	18
3.2.2.	Variabel	19
3.3.	Pelabelan dan Validasi Data	19
3.4.	Preprocessing data	20
3.4.1.	Cleansing Data	21
3.4.2.	Case Folding	21
3.4.3.	Stopword removal	22
3.4.4.	Tokenizing	22
3.5.	Klasifikasi menggunakan Random Forest Clasifier	23
3.5.1.	TF-IDF	23
3.5.2.	Implementasi	27
3.5.3.	Testing Confusion Matrix	28
3.5.4.	Pengujian Performa Algoritma	31
BAB IV		32
HASIL DAN PEMBAHASAN		32
4.1.	Implementasi Sistem	32
4.1.1.	Kebutuhan Perangkat Keras dan Perangkat Lunak	32
4.1.2.	Data Preprocessing	32
4.1.2.1.	Cleaning Data	34

4.1.2.2.	Case Folding	35
4.1.2.3.	Stopword Removal	35
4.1.2.4.	Tokenizing	35
4.1.3.	TF-IDF	36
4.1.4.	Training	37
4.1.5.	Testing	37
4.2.	Perbandingan Antar Metode.....	38
4.2.1.	Metode <i>Multinomial Naive Bayes</i> (MNB).....	38
4.2.2.	Metode <i>Decision Tree Classifier</i>	39
4.2.3.	Metode <i>KNeighbors Classifier</i> (KNN)	40
BAB V	44
PENUTUP	44
5.1.	KESIMPULAN	44
5.2.	SARAN.....	44
DAFTAR PUSTAKA	46



DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	8
Tabel 2.2 Confusion Matrix	15
Tabel 3.1 Variabel	19
Tabel 3.4 Case Folding	21
Tabel 3.5 Tokenizing	23
Tabel 3.6 Ilustrasi Sampel Data pada Perhitungan TF.....	25
Tabel 3.7 Ilustrasi Sampel Data pada Perhitungan DF	25
Tabel 3.8 Ilustrasi Sampel Data pada Perhitungan IDF.....	26
Tabel 3.9 Perhitungan TF.....	27
Tabel 3.10 Confusion Matrix	28
Tabel 3.11 Hitung Precission	28
Tabel 3.12 Hitung Presisi	29
Tabel 3.13 Recall	29
Tabel 3.14 Hitung Recall	30
Tabel 4.15 Perbandingan metode.....	40

DAFTAR GAMBAR

Gambar 2.1 Contoh Visualisasi Word Clouds	16
Gambar 3.2 Dataset.....	19
Gambar 3.3 Preprocessing data.....	20
Gambar 4.1 Import Data	33
Gambar 4.5 Case Folding.....	35
Gambar 4.6 Stopword Removal.....	35
Gambar 4.7 Tokenizing.....	36
Gambar 4.8 Hasil Cleaning Text.....	36
Gambar 4.9 TF-IDF	37
Gambar 4.10 Training Data	37
Gambar 4.11 Testing.....	38
Gambar 4.12 Testing Metode MNB.....	39
Gambar 4.13 Testing Metode Decision Tree	39
Gambar 4.14 Testing Metode KNN	40
Gambar 4.15 Perbandingan Metode.....	41
Gambar 4.16 Perbandingan Presisi	41
Gambar 4.17 Perbandingan Recall.....	42
Gambar 4.18 Perbandingan Akurasi	42
Gambar 4.19 Perbandingan F1-Score	43

INTISARI

Kegiatan masyarakat pada saat ini tidak terlepas dari media sosial. Media sosial yang sedang banyak diminati oleh khalayak saat ini adalah Instagram. Di dalam Instagram ini berisikan kumpulan foto maupun video dan caption berupa text. Dimana caption ini digunakan untuk mengungkapkan tentang diri mereka sendiri dan hal yang sedang dirasakannya atau menggambarkan tentang diri mereka sendiri, maka dari data yang ada dapat dimanfaatkan untuk bidang psikologi yaitu menganalisis kepribadian seseorang secara digital.

Dengan banyaknya pengguna Instagram dengan caption-caption yang ada maka dapat digunakan sebagai data untuk menganalisis dari kepribadian seseorang yang dinilai dengan kepribadian DISC. Kepribadian DISC itu sendiri merupakan sebuah klasifikasi yang membedakan tipe kepribadian seseorang menjadi empat tipe yaitu tipe kepribadian Dominance, Influence, Steadiness, dan Compliance.

Pada penelitian ini menggunakan algoritma Random Forest Classifier dengan memperoleh akurasi sebesar 59,8% dan performa yang dihasilkan metode Random Forest lebih baik dibandingkan dengan metode *Multinomial Naïve Bayes* (MNB), Decision Tree, dan *K-Nearest Neighbor* (KNN).

Kata kunci : *Instagram, DISC, Random Forest Classifier*

ABSTRACT

Community activities at this time cannot be separated from social media. The most popular social media nowadays is Instagram. In this Instagram contains a collection of photos and videos and captions in the form of text. Where this caption is used to express about themselves and what they are feeling or describe themselves, then the existing data can be used for the field of psychology, namely analyzing a person's personality digitally.

With so many Instagram users with existing captions, it can be used as data to analyze the personality of a person who is judged by the DISC personality. DISC personality itself is a classification that distinguishes a person's personality type into four types, namely the Dominance, Influence, Steadiness, and Compliance personality types.

In this study using the Random Forest Classifier algorithm with an accuracy of 59.8% and the resulting performance of the Random Forest method is better than the Multinomial Naïve Bayes (MNB), Decision Tree, and K-Nearest Neighbor (KNN) methods.

Keywords: *Instagram, DISC, Random Forest Classifier*