

**ANALISIS PENCARIAN PROFIL JOBSEEKER MENGGUNAKAN DATA
TWITTER DENGAN ALGORITMA NAIVE BAYES CLASSIFIER**

SKRIPSI



Disusun oleh

Harish Setyo Hudnanto

16.11.0048

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

2019

**ANALISIS PENCARIAN PROFIL JOBSEEKER MENGGUNAKAN DATA
TWITTER DENGAN ALGORITMA NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Teknik Informatika



disusun oleh

Harish Setyo Hudnanto

16.11.0048

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

2019

PERSETUJUAN

SKRIPSI

ANALISIS PENCARIAN PROFIL JOBSEEKER MENGGUNAKAN DATA TWITTER DENGAN ALGORITMA NAIVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Harish Setyo Hudnanto

16.11.0048

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Agustus 2019

Dosen Pembimbing,



Anggit Dwi Hartanto, M.Kom

NIK. 190302163

PENGESAHAN

SKRIPSI

ANALISIS PENCARIAN PROFIL JOBSEEKER MENGGUNAKAN DATA TWITTER DENGAN ALGORITMA NAIVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Harish Setyo Hudnanto

16.11.0048

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Oktober 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ahlihi Masruro, M.Kom
NIK. 190302148



Anggit Dwi Hartanto, M.Kom
NIK. 190302163



Windha Mega Pradnya D, M.Kom
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 25 Oktober 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, 25 Oktober 2019



Harish Setyo Hudnanto

NIM. 16.11.0048

MOTTO

“Apa artinya kaki bila kau tak berjalan
Apa guna mata bila tak menatap masa depan
Untuk apa bermimpi bila kau tak melangkah
Untuk apa kesempatan bila tak mengambil celah”
(Bondan Prakoso – Xpresikan)

“Kau masih terlalu muda untuk memikirkan kegagalan”
(Silver Fang – One Punch Man)

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

“Tidak menyerah adalah sihir terkuatku !”
(Asta – Black Clover)

“Lahir, sekolah, bekerja, mati
Sistem hidupku berpatok pada materi”
(Bondan Prakoso – S.O.S)

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 Didik Purwantara dan Ibu Tini Sutiasih. 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. Mbah putri, yang telah memberikan tempat bagi penulis untuk bisa tidur dengan nyenyak selama menempuh pendidikan di Amikom Yogyakarta.
3. Adikku tersayang, Cahya Rizqi Amanah, yang sudah memberikan dukungan dan doa kepada penulis.
4. Teman-teman sekelas, 16-S1 IF-01 yang mohon maaf tidak bisa penulis sebutkan satu persatu karena terlalu banyak.
5. Sahabat pejuang skripsi kontrakan Bos Pared dan kontrakan Gelatik : Farid, Andi, Hananto, Ariel, Taufiq, Arik, Arif Marda, Arif Setyo, Indra, Ardi Bayu, Dian Noviard, Wahyudi, dan Irfan. Terima kasih untuk dukungan, doa, tempat, dan waktu yang sudah kita habiskan bersama. Semoga kita semua menjadi orang-orang yang beruntung.
6. Sahabat seperjuangan, Aji Kamaludin, Aminudin Noor Ichsan, dan Maful Prayoga Arnandi. Terimakasih sudah berjuang bersama, saling bertukar pemikiran, dan berkeluh kesah bersama. Semoga kita tidak pernah melupakan kenangan perjuangan ini.

KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT atas segala limpahan ridho, hidayah, serta inayah-Nya sehingga penelitian dengan judul “Analisis Pencarian Profil Jobseeker Menggunakan Data Twitter Dengan Algoritma Naive Bayes Classifier” 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, Bapak Anggit Dwi Hartanto, M.Kom, Ibu Sumarni Adi, S.Kom., M.Cs, dan Prof. Dr. Ema Utami, S.Si., M.Kom.
- Kepada sahabat dan teman-teman yang telah memberikan dukungan serta doa yang tulus kepada penulis.
- Kepada kampus tercinta, Universitas Amikom Yogyakarta.

Yogyakarta, 24 Oktober 2019

Harish Setyo Hudnanto

DAFTAR ISI

JUDUL	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN.....	IV
MOTTO	V
PERSEMBAHAN.....	VI
KATA PENGANTAR	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL.....	XI
DAFTAR GAMBAR	XII
INTISARI.....	XIII
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Maksud dan Tujuan Penelitian	5
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.6.1 Metode Pengumpulan Data.....	6
1.6.2 Metode Analisis	6
1.6.3 Metode Perancangan.....	6
1.6.4 Implementasi.....	6
1.6.5 Metode Pembahasan.....	6
1.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	8
2.1 Perumusan Masalah.....	8
2.2 Landasan Teori.....	11
2.2.1 Pengertian Twitter.....	11
2.2.2 Pengertian Karakter.....	12

2.2.3	Teori DISC	12
2.2.4	Text Mining.....	13
2.2.5	Pengertian Web Scraping.....	14
2.2.6	Pengertian Klasifikasi	14
2.2.7	Klasifikasi Teks.....	15
2.2.8	Tokenisasi	15
2.2.9	<i>Text Transformation</i> dan Pembobotan W-IDF	15
2.2.10	Naive Bayes Classifier	16
2.2.11	Confusion Matrix	17
BAB III ANALISIS DAN PERANCANGAN		19
3.1	Analisis Sistem	19
3.1.1	Analisis Kebutuhan Fungsional	19
3.1.2	Analisis Kebutuhan Non-Fungsional	19
3.2	Gambaran Umum Sistem	20
3.3	Analisis Algoritma	21
3.3.1	Pengumpulan Data	21
3.3.2	Validasi Data Set.....	22
3.3.3	Pre-processing.....	23
3.3.3.1	Case Folding.....	24
3.3.3.2	<i>Tweet Cleaning</i>	25
3.3.3.3	Tokenizing.....	26
3.3.4	Seleksi Data.....	28
3.3.5	Pembobotan W-IDF	29
3.3.5.1	Menghitung TF pada <i>Tweet</i>	30
3.3.5.2	Menghitung TF kata pada semua <i>Tweet</i>	31
3.3.5.3	Menghitung nilai W-IDF.....	32
3.3.5.4	Pengelompokkan nilai W-IDF berdasarkan label	33
3.3.6	Naïve Bayes Classifier	34
3.3.7	Confusion Matrix	35
3.3.7.1	Validasi Data Set.....	35
3.3.7.2	Hasil Olah Data Menggunakan <i>Naive Bayes Classifier</i>	37

3.3.7.3	Hasil Evaluasi.....	38
3.4	Analisis Kebutuhan Data.....	45
3.4.1	Perancangan Data Set.....	45
3.4.2	Perancangan Data Training.....	46
3.5	Perancangan Antarmuka.....	46
3.5.1	Perancangan Halaman <i>Scraping</i>	46
3.5.2	Perancangan Halaman Detail Akun.....	47
3.5.3	Perancangan Halaman <i>Preprocessing</i>	48
3.5.4	Perancangan Halaman Detail <i>Preprocessing</i> Akun.....	49
3.5.5	Perancangan Halaman Klasifikasi.....	50
3.5.6	Perancangan Halaman Hasil Klasifikasi.....	50
BAB IV	IMPLEMENTASI DAN PEMBAHASAN	52
4.1	Deskripsi Implementasi.....	52
4.2	Implementasi Pengumpulan Data Tweet.....	52
4.3	Implementasi Proses <i>Preprocessing</i>	54
4.4	Implementasi Pembobotan dan Klasifikasi.....	57
4.5	Evaluasi dan Validasi.....	59
4.6	Implementasi Antarmuka dan Panduan Penggunaan Sistem.....	60
4.6.1	Halaman Masuk.....	60
4.6.2	Halaman Akun <i>Twitter</i>	61
4.6.3	Halaman <i>Crawling</i> Data Akun <i>Twitter</i>	62
4.6.4	Halaman <i>Preprocessing</i>	63
4.6.5	Halaman Klasifikasi Karakter.....	64
4.6.6	Halaman Evaluasi.....	66
4.7	Pengujian Sistem.....	67
4.7.1	White Box Testing.....	67
4.7.2	Black Box Testing.....	68
BAB V	PENUTUP	70
5.1	Kesimpulan.....	70
5.2	Saran.....	71
DAFTAR PUSTAKA	72

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian.....	10
Tabel 2.2 Confusion Matrix	17
Tabel 3.1 Contoh Data Tweet	21
Tabel 3.2 Contoh Pelabelan Kata.....	22
Tabel 3.3 Hasil Case Folding	24
Tabel 3.4 Hasil Tweet Cleaning	26
Tabel 3.5 Hasil Tokenizing	27
Tabel 3.6 Hasil Seleksi Data	28
Tabel 3.7 Perhitungan TF pada setiap Tweet.....	31
Tabel 3.8 TF Kata pada Semua Tweet	31
Tabel 3.9 Perhitungan Nilai WIDF	32
Tabel 3.10 Nilai WIDF berdasarkan label	33
Tabel 3.11 Hasil Klasifikasi	34
Tabel 3.12 Contoh Data Set Akun Twitter Personal.....	35
Tabel 3.13 Contoh Data Set Kata Validasi	36
Tabel 3.14 Validasi Data Set.....	37
Tabel 3.15 Hasil Klasifikasi Akun Twitter	37
Tabel 3.16 Perbandingan Data Set Akun dan Hasil Klasifikasi.....	38
Tabel 3.17 Confusion Matrix	44
Tabel 3.18 Precision dan Recall.....	45
Tabel 4.1 Black Box Testing.....	68

DAFTAR GAMBAR

Gambar 3.1 Alur proses pada sistem.....	21
Gambar 3.2 Proses Preprocessing	23
Gambar 3.3 Proses Case Folding	24
Gambar 3.4 Proses Tweet Cleaning.....	25
Gambar 3.5 Proses Tokenizing	27
Gambar 3.6 Proses Pembobotan W-IDF	30
Gambar 3.7 Data Set Akun, Tweet, Label	46
Gambar 3.8 Perancangan Halaman Scraping.....	47
Gambar 3.9 Perancangan Halaman Detail Akun	48
Gambar 3.10 Perancangan Halaman Preprocessing.....	49
Gambar 3.11 Perancangan Halaman Detail Preprocessing Akun.....	49
Gambar 3.12 Perancangan Halaman Klasifikasi.....	50
Gambar 3.13 Perancangan Halaman Hasil Klasifikasi	51
Gambar 3.14 Perancangan Halaman Evaluasi	51
Gambar 4.1 Implementasi Confusion Matrix	60
Gambar 4.2 Tampilan Halaman Masuk	61
Gambar 4.3 Tampilan Halaman Akun Twitter	62
Gambar 4.4 Tampilan Halaman Crawling Data.....	63
Gambar 4.5 Tampilan Halaman Preprocessing.....	64
Gambar 4.6 Tampilan Halaman Detail Preprocessing.....	64
Gambar 4.7 Tampilan Halaman Klasifikasi Karakter.....	65
Gambar 4.8 Tampilan Halaman Detail Klasifikasi Karakter.....	65
Gambar 4.9 Tampilan Tahapan Pemrosesan Evaluasi.....	66
Gambar 4.10 Tampilan Halaman Evaluasi	67

INTISARI

Ketika mendaftar lowongan pekerjaan, terdapat beberapa tahap yang harus ditempuh oleh pelamar kerja atau *Jobseeker*. Salah satunya adalah tahap wawancara. Wawancara biasanya dilakukan oleh seorang *Human Resources* atau HR untuk mengetahui lebih jauh karakter serta kepribadian pelamar kerja tersebut. Karena terbatasnya informasi tentang pelamar kerja tersebut, maka akan cukup sulit bagi seorang HR untuk mengajukan pertanyaan sehingga didapatkan informasi yang cocok untuk kebutuhan perusahaan.

Twitter sebagai salah satu platform media sosial yang sangat populer di Indonesia menjadikan penggunanya aktif dalam menuliskan pemikiran-pemikiran, perasaan, kritikan, bahkan kehidupan pribadi kedalam suatu cuitan atau yang biasa disebut dengan *Tweet*. Dengan adanya media sosial ini, diharapkan akan membantu pekerjaan seorang HR dalam memperoleh informasi berupa karakter atau kepribadian dari pelamar kerja tersebut. Maka dari itu, dibuatlah sistem untuk menganalisis karakter kepribadian seseorang berdasarkan teori kepribadian DISC dan Algoritma Naive Bayes Classifier yang diolah berdasarkan cuitan-cuitan dari akun *Twitter* pribadi pelamar kerja.

Pengujian sistem dilakukan menggunakan 120 akun *Twitter* yang sebelumnya telah divalidasi secara manual oleh seorang pakar kepribadian atau dalam hal ini seorang psikolog yang dijadikan Data Set. Hasil evaluasi yang didapat dari penelitian ini, membuktikan bahwa sistem dengan Algoritma Naive Bayes Classifier ini telah berhasil mengklasifikasikan karakter atau kepribadian seseorang berdasarkan cuitan-cuitan dari akun *Twitter* pribadi pelamar kerja dengan akurasi 36,67 %.

Kata kunci : Analisis karakter, Twitter, teori DISC, Algoritma, Naive Bayes Classifier, klasifikasi, Jobseeker.

ABSTRACT

When registering a job opening, there are several stages that must be taken by job applicants or Jobseeker. One of them is the interview stage. Interviews are usually conducted by a Human Resources or HR to find out more about the character and personality of the job applicant. Because of the limited information about job applicants, it will be quite difficult for a HR to ask questions so that the information obtained is suitable for the company's needs.

Twitter as one of the most popular social media platforms in Indonesia makes its users active in writing thoughts, feelings, criticisms, even personal life into a chat or what is commonly called a Tweet. With this social media, it is expected to help the work of an HR in obtaining information in the form of character or personality from the job applicant. Therefore, a system was made to analyze the character or personality of a person based on the DISC personality theory and the Naive Bayes Classifier Algorithm which was processed based on tweets from the applicant's personal Twitter account.

The system testing was carried out using 120 Twitter accounts that had previously been manually validated by a personality expert or in this case a psychologist who was made a Data Set. Evaluation results obtained from this study, prove that the system with the Naive Bayes Classifier Algorithm has succeeded in classifying a person's character or personality based on tweets from job applicants' personal Twitter accounts with an accuracy of 36.67%.

Keyword : *Character Analysis, Twitter, DISC theory, Algorithm, Naive Bayes Classifier, classification, Jobseeker.*