

TUGAS AKHIR

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR (KNN)* DAN *ROCCHIO*
DALAM KLASIFIKASI TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER
STUDI KASUS: FAKULTAS TEKNIK**



**FARID ACHMAD ARIF ADANI
1610651055**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2020

TUGAS AKHIR

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR (KNN)* DAN *ROCCHIO*
DALAM KLASIFIKASI TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER
STUDI KASUS: FAKULTAS TEKNIK**

**Disusun untuk Melengkapi dan Memenuhi Syarat Kelulusan Progam Strata 1
Progam Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember**



FARID ACHMAD ARIF ADANI

1610651055

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2020

HALAMAN PERSETUJUAN

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR (KNN)* DAN *ROCCHIO*
DALAM KLASIFIKASI TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER
STUDI KASUS: FAKULTAS TEKNIK**

Oleh:

Farid Achmad Arif Adani

1610651055

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

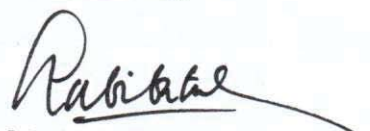
Universitas Muhammadiyah Jember

Disetujui oleh:

Pembimbing I


Deni Arifianto, M.Kom.
NIDN. 0718068103

Pembimbing II


Habibatul Azizah Al Faruq, M.Pd.
NIDN. 0718128901

HALAMAN PENGESAHAN

**PERBANDINGAN ALGORITMA *K-NEAREST NEIGHBOR (KNN)* DAN *ROCCHIO*
DALAM KLASIFIKASI TUGAS AKHIR
UNIVERSITAS MUHAMMADIYAH JEMBER
STUDI KASUS: FAKULTAS TEKNIK**

Oleh:

**Farid Arif Arif Adani
1610651055**

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir tanggal
27 Juni 2020 sebagai salah satu syarat kelulusan dan mendapatkan gelar
Sarjana Komputer (S.Kom)
di

Universitas Muhammadiyah Jember

Disetujui oleh:

**Dosen Penguji:
Penguji I**



Wiwik Suharso, S.Kom., M.Kom.
NIDN. 0006097601

Penguji II

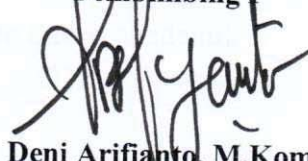


Ginanjar Abdurrahman, S.Si., M.Pd.
NIDN. 0714078704

**Mengesahkan,
Dekan Fakultas Teknik**


Nanang Saiful Rizal, S.T., M.T.
NIDN. 0705047806

**Dosen Pembimbing:
Pembimbing I**



Deni Arifianto, M.Kom.
NIDN. 0718068103

Pembimbing II



Habibatul Azizah Al Faruq, M.Pd.
NIDN. 0718128901

**Mengetahui,
Ketua Program Studi Teknik Informatika**


Triawan Adi Cahyanto, M.Kom.
NIDN. 0702098804

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

NAMA : FARID ACHMAD ARIF ADANI
NIM : 16 1065 1055
INSTITUSI : Strata-1 Program Studi Teknik Informatika Fakultas
Teknik, Universitas Muhammadiyah Jember.

Menyatakan bahwa Tugas Akhir yang berjudul “**Perbandingan Algoritma *K-Nearest Neighbor (KNN)* dan *Rocchio* Dalam Klasifikasi Tugas Akhir Universitas Muhammadiyah Jember Studi Kasus: Fakultas Teknik**” bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi akademik.

Jember, 27 Juni 2020

Yang Menyatakan,



Farid achmad arif adani

NIM. 16 1065 1055

KATA PENGANTAR

Alhamdulillah segala puji bagi Allah yang senantiasa memberikan nikmat yang sangat luar biasa dalam hidup ini kepada penulis berupa nikmat untuk selalu bersujud dan meminta ampunan kepada-Nya. Serta berkat rahmat-Nya penulis diberikan kemudahan dalam menyelesaikan studi di kampus Universitas Muhammadiyah Jember.

Atas segala upaya, bimbingan dan arahan dari semua pihak, penulis mengucapkan banyak terima kasih. Dalam kesempatan ini penulis mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Allah SWT atas segala rahmat, nikmat dan hidayah-Nya, terima kasih atas kehendak-Mu telah memberikan kesempatan ini untuk mendapatkan gelar Sarjana Komputer.
2. Bapak Nanang Saiful Rizal, S.T., M.T., selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Triawan Adi Cahyanto, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto, M.Kom., selaku dosen pembimbing I dan Ibu Habibatul Azizah Al Faruq, M.Pd., selaku dosen pembimbing II yang telah meluangkan waktu untuk memberi arahan dan membimbing penulis dalam menyelesaikan Tugas Akhir ini.
5. Ibu Rosita Yanuarti, S.Kom., M.Cs., yang telah membantu untuk memberi arahan dan membimbing penulis dalam menyelesaikan Tugas Akhir ini.
6. Bapak Wiwik Suharso, S.Kom., M.Kom., selaku dosen penguji 1 dan Bapak Ginanjar Abdurrahman, S.Si., M.Pd., selaku dosen penguji II yang telah memberikan saran dan masukan dalam penyelesaian Tugas Akhir ini.
7. Kepada Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmunya kepada penulis.
8. Kepada Bapak dan Ibu Staf dan Laboran Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah membantu penulis dalam melaksanakan tugas akhir.
9. Kepada orang tua penulis Bapak Suparto dan Ibu Farida yang selalu mendoakan, memberikan motivasi, dukungan dan memberikan kasih sayang sehingga penulis dapat menyelesaikan tugas akhir ini.

10. Kepada Himajo yang selalu memberi semangat, bantuan, dan nasihat ke penulis untuk menyelesaikan tugas akhir ini.
11. Kepada Pemburu Airdrop, Yuka, Feni dan Hamdhan yang telah membantu untuk memberikan semangat, dukungan, bantuan, nasihat, saran, dan doa ke penulis untuk menyelesaikan tugas akhir ini.
12. Serta kepada teman-teman Program Studi Teknik Informatika Angkatan 2016 yang telah memberikan doa, dukungan, dan bantuannya.

Dan untuk semuanya, semoga Allah menunjukkan kepada kita pada jalan yang telah diridhoi oleh-Nya, diberikan ilmu yang barokah, serta diampuni dosa- dosa kita, dan senantiasa tercurah limpahkan rahmat, taufiq dan hidayah-Nya kepada kita semua.

Aamiin... Yaa Robbal Alamiin...

Jember, 27 Juni 2020

Penulis

MOTTO

Knowledge And Faith Are Always My Companions

“Pengetahuan dan keyakinan selalu menjadi temanku”

Rafaela, Mobile Legend

A Painless Lesson Is Also Meaningless

“Pelajaran tanpa rasa sakit juga tidak berarti”

Lolita, Mobile Legend

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
MOTTO	ix
DAFTAR ISI	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR GRAFIK	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 <i>Text Mining</i>	5
2.2 <i>Text Preprocessing</i>	5
2.3 <i>Term Weighting TF - IDF</i>	6
2.4 <i>K-Nearest Neighbor</i>	6
2.5 <i>Rocchio</i>	7
2.6 <i>K-Fold Cross Validation</i>	8
2.7 <i>Confusion Matrix</i>	8
2.8 Akurasi, Presisi, dan <i>Recall</i>	9
2.9 <i>Jupyter Notebook</i>	10

BAB III METODE PENELITIAN	11
3.1 Rancangan Penelitian	11
3.2 Pengumpulan Data	11
3.3 <i>Pre-processing</i>	12
3.4 Pembobotan	14
3.5 Klasifikasi.....	21
3.5.1 Metode <i>K-Nearest Neighbor</i>	21
3.5.2 Metode <i>Rocchio</i>	22
3.6 Validasi dan Evaluasi	28
BAB IV PEMBAHASAN.....	29
4.1 Data.....	29
4.2 Hasil Klasifikasi	29
4.2.1 Pengujian Menggunakan <i>Kfold 2</i>	29
4.2.2 Pengujian Menggunakan <i>Kfold 3</i>	33
4.2.3 Pengujian Menggunakan <i>Kfold 5</i>	37
4.2.4 Pengujian Menggunakan <i>Kfold 6</i>	41
4.2.5 Pengujian Menggunakan <i>Kfold 10</i>	45
4.3 Partisi Data Abstrak dan Bab 1	50
BAB V KESIMPULAN DAN SARAN	51
5.1 Kesimpulan.....	54
5.2 Saran	55
DAFTAR PUSTAKA	56
LAMPIRAN.....	59

DAFTAR GAMBAR

Gambar 3.1 Alur Metode Penelitian	11
Gambar 4.1 Tree Data Tugas Akhir.....	51

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	8
Tabel 2.2 Rumus Akurasi, Presisi, dan <i>Recall</i>	9
Tabel 3.1 Hasil <i>Case Folding</i>	12
Tabel 3.2 Hasil <i>Stop Character Removal</i>	13
Tabel 3.3 Hasil <i>Token</i>	13
Tabel 3.4 Hasil <i>Stopwordremoval</i>	14
Tabel 3.5 Hasil <i>Stemming</i>	14
Tabel 3.1 Perhitungan <i>Term Frequency</i>	15
Tabel 3.2 Perhitungan <i>N/dfti</i>	17
Tabel 3.3 Perhitungan IDF	18
Tabel 3.4 Nilai Bobot <i>Term</i> pada Tiap Dokumen.	19
Tabel 3.5 <i>Rank</i> Kedekatan Jarak <i>Euclidean</i>	22
Tabel 3.6 Hasil Perhitungan <i>Centroid</i>	22
Tabel 3.7 Hasil Perhitungan Vektor <i>Centroid</i> x Vektor D6.....	24
Tabel 3.8 Hasil Panjang Vektor Tiap Dokumen	25
Tabel 3.9 Hasil Panjang Vektor D6.....	26
Tabel 3.10 Hasil Perhitungan $v \rightarrow d1$. $v \rightarrow d2$ Masing – Masing <i>Centroid</i>	28
Tabel 3.11 Hasil <i>Cosine Similarity</i>	28
Tabel 4.1 Hasil Pencarian Nilai K pada <i>Kfold 2</i>	29
Tabel 4.2 <i>Confusion Matrix</i> pada Algoritma <i>KNN Kfold 2</i>	31
Tabel 4.3 <i>Confusion Matrix</i> pada algoritma <i>Rocchio Kfold 2</i>	32
Tabel 4.4 Hasil Pencarian Nilai K pada <i>Kfold 3</i>	33
Tabel 4.5 <i>Confusion Matrix</i> pada Algoritma <i>KNN Kfold 3</i>	35
Tabel 4.6 <i>Confusion Matrix</i> pada Algoritma <i>Rocchio Kfold 3</i>	36
Tabel 4.7 Hasil Pencarian Nilai K pada <i>Kfold 5</i>	37
Tabel 4.8 <i>Confusion Matrix</i> pada Algoritma <i>KNN Kfold 5</i>	39
Tabel 4.9 <i>Confusion Matrix</i> pada Algoritma <i>Rocchio Kfold 5</i>	40
Tabel 4.10 Hasil Pencarian Nilai K pada <i>Kfold 6</i>	42
Tabel 4.11 <i>Confusion Matrix</i> pada Algoritma <i>KNN Kfold 6</i>	43

Tabel 4.12 <i>Confusion Matrix</i> pada Algoritma <i>Rocchio Kfold 6</i>	44
Tabel 4.13 Hasil Pencarian Nilai K pada <i>Kfold 10</i>	46
Tabel 4.14 <i>Confusion Matrix</i> pada Algoritma <i>KNN Kfold 10</i>	47
Tabel 4.15 <i>Confusion Matrix</i> pada Algoritma <i>Rocchio Kfold 10</i>	48
Tabel 4.16 Jarak <i>Euclidean KNN</i> Abstrak dan Bab1 Pada Skenario 1	50
Tabel 4.17 Jarak <i>Cosine</i> pada <i>Rocchio</i> Abstrak dan Bab1 pada Skenario 1	51
Tabel 4.18 Jarak <i>Euclidean KNN</i> Abstrak dan Bab1 pada Skenario 2.....	52
Tabel 4.19 Jarak <i>Cosine</i> pada <i>Rocchio</i> Abstrak dan Bab1 pada Skenario 2	52

DAFTAR GRAFIK

Grafik 4.1 Hasil Pengujian Pada <i>Kfold 2</i>	30
Grafik 4.2 Hasil Pengujian Pada <i>Kfold 3</i>	34
Grafik 4.3 Hasil Pengujian Pada <i>Kfold 5</i>	38
Grafik 4.4 Hasil Pengujian Pada <i>Kfold 6</i>	42
Grafik 4.5 Hasil Pengujian Pada <i>Kfold 10</i>	46