

**IMPLEMENTASI ALGORITMA K-NN UNTUK KLASIFIKASI BIDANG
DOSEN BERDASARKAN JUDUL SKRIPSI**

SKRIPSI



disusun oleh

Nurrofiqi Ankisqlantari

18.11.2310

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**IMPLEMENTASI ALGORITMA K-NN UNTUK KLASIFIKASI BIDANG
DOSEN BERDASARKAN JUDUL SKRIPSI**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Nurroflqi Ankisqiantari

18.11.2310

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA K-NN UNTUK KLASIFIKASI BIDANG DOSEN BERDASARKAN JUDUL SKRIPSI

yang dipersiapkan dan disusun oleh

Nurrofiq Anksiqantari

18.11.2310

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Maret 2022

Dosen Pembimbing,

Windha Mega Pradnya D, M.Kom
NIK. 190302185

PENGESAHAN

SKRIPSI

**IMPLEMENTASI ALGORITMA K-NN UNTUK KLASIFIKASI BIDANG
DOSEN BERDASARKAN JUDUL SKRIPSI**

yang dipersiapkan dan disusun oleh

Nurroflqi Anksqlantari

18.11.2310

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Nuraini, M.Kom
NIK. 190302066

Yuli Astuti, M.Kom
NIK. 190302146

Windha Mega Pradnya D, M.Kom
NIK. 190302185

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 02 Maret 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, 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 etika dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 05 Februari 2022




Nurrofiq Aakiesqattari

NIM. 18.11.2310

MOTTO

لَا تَخْزَنُ إِنَّ اللَّهَ مَعِنَ

“Janganlah susah, Allah selalu bersamamu.”

~At Taubah ayat 40

Andalkan Allah disetiap hidupmu, Allah selalu bersama kita.



PERSEMBAHAN

Alhamdulillah robbil' alamin segala puji bagi Allah SWT yang telah memberikan kelancaran dan kebarokahan bagi penulis dalam menyusun skripsi ini, sehingga dapat selesai dengan baik dan lancar. Dengan ini saya persembahkan skripsi ini kepada semua pihak yang membantu support perkuliahan hingga menyelesaikan studi untuk mendapat gelar sarjana, yaitu :

1. Kedua orang tua dan adik-adik penulis yang selalu mendoakan, menasehati, mendukung dengan cinta kasih sayang.
2. Keluarga besar penulis yang selalu menyertakan doanya
3. Almuzhidul Mujid, dan Adi Zaenul yang selalu membantu dan memberi masukan dalam pengerjaan program skripsi ini.
4. Lasmita Triana, Agita Putri K. , Muh. Farhan K.D, teman-teman seangkatan, sekelas, sedosen bimbingan, yang saling membantu, berproses bersama sampai pendadaran bersama, lulus bersama.
5. Sahabat-sahabat penulis, Erwinda Latifah Hanum, Lasmita Triana, Emilia Tsiompah, Fanani Amalia H.J.F, Elyasi Royana Firdaus yang selalu menyemangati, mendukung, dan memberikan solusi.
6. Teman – teman SMP yang mensupport dan memberikan kesan baik untuk penulis.
7. Teman – teman Angkatan 18 Informatika 7, khususnya untuk teman-teman yang pernah satu kelompok atau project-an, yang menyelesaikan tugas bersama sama dan melancarkan tugas ini.
8. Dan semua pihak yang disekitar penulis, terimakasih banyak atas dukungannya. Semoga sehat selalu.

KATA PENGANTAR

Puji syukur kepada Allah SWT yang telah memberikan kenikmatan, hidayah, kemudahan dan kelancaran, shalawat serta salam senantiasa penulis panjatkan kepada Nabi Muhammad SAW, sehingga penulis senantiasa diberikan kelancaran dan kebarokahan dalam menyelesaikan skripsi dengan judul “Implementasi Algoritma K-NN untuk Klasifikasi Bidang Dosen berdasarkan Judul Skripsi” dengan baik.

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, oleh karena itu pada kesempatan kali ini penulis menyampaikan rasa hormat dan terima kasih kepada :

1. Bapak M. Suyanto, Prof., Dr., M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta dan Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Program Studi Informatika.
3. Ibu Windha Mega Pradnya D, M.Kom selaku dosen pembimbing yang telah memberikan pengarahan penulis serta telah sabar membimbing dalam penyusunan skripsi ini.
4. Ibu Nuraini, M.Kom dan Ibu Yuli Astuti, M.Kom selaku dosen penguji. Terima kasih saran yang telah diberikan selama pengujian untuk memperbaiki penelitian menjadi lebih baik lagi.

Penulis menyadari skripsi ini masih terdapat kekurangan, maka dari itu perlu kritik dan saran yang bersifat membangun dari berbagai pihak akan diterima dengan lapang dada untuk memperbaiki karya-karya selanjutnya. Semoga skripsi ini dapat bermanfaat bagi penulis dan pembaca.

DAFTAR ISI

COVER	1
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
<i>abstract</i>	xiv
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	2
1.3 BATASAN MASALAH	2
1.4 MAKSUD DAN TUJUAN PENELITIAN	3
1.5 MANFAAT PENELITIAN	3
1.6 METODE PENELITIAN	3
1.6.1 Metode Pengumpulan Data	3
1.6.2 Metode Analisis	4
1.6.3 Metode Perancangan	4
1.6.4 Metode Implementasi dan Testing	4
1.7 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI	6

2.1	KAJIAN PUSTAKA.....	6
2.2	DASAR TEORI.....	12
2.2.1	<i>Data Science</i>	12
2.2.2	<i>Machine Learning</i>	12
2.2.3	<i>Text Mining</i>	12
2.2.4	<i>Text Preprocessing</i>	13
2.2.5	<i>TF-IDF</i>	13
2.2.6	<i>K-NN</i>	14
2.2.7	<i>Confusion Matrix</i>	16
2.2.8	<i>Black-box</i>	17
2.2.9	<i>Python</i>	18
2.2.10	<i>Jupyter Notebook</i>	18
2.2.11	<i>Website</i>	18
2.2.12	<i>Flask</i>	19
BAB III ANALISIS DAN PERANCANGAN.....		20
3.1	GAMBARAN UMUM.....	20
3.2	ALAT DAN BAHAN PENELITIAN.....	21
3.2.1	<i>Hardware</i>	21
3.2.2	<i>Software</i>	22
3.3	ALUR PENELITIAN.....	22
3.3.1	<i>Pengambilan Data</i>	24
3.3.2	<i>Text Preprocessing</i>	32
3.3.3	<i>Data Split</i>	39
3.3.4	<i>TF-IDF</i>	39
3.3.5	<i>Klasifikasi dengan K-Nearest Neighbor (K-NN)</i>	50
3.3.6	<i>Confusion Matrix</i>	57
3.4	PERANCANGAN ANTAR MUKA (INTERFACE).....	58
BAB IV IMPLEMENTASI DAN PEMBAHASAN.....		59
4.1	DESKRIPSI IMPLEMENTASI.....	59
4.2	IMPLEMENTASI MEMASUKKAN DATA.....	59



4.3	IMPLEMENTASI PELABELAN DATA.....	61
4.4	IMPLEMENTASI PERANCANGAN FUNGSI DAN SISTEM.....	62
4.4.1	<i>Implementasi Fungsi Text Preprocessing</i>	62
4.4.2	<i>Implementasi Data Split</i>	64
4.4.3	<i>Implementasi TF-IDF</i>	65
4.4.4	<i>Implementasi Algoritma K-NN</i>	66
4.4.5	<i>Implementasi Confusion Matrix</i>	66
4.5	IMPLEMENTASI PERBANDINGAN PENGUJIAN.....	67
4.6	IMPLEMENTASI INTERFACE.....	68
4.6.1	<i>Halaman Utama / Beranda</i>	68
4.6.2	<i>Input Judul Skripsi</i>	69
4.6.3	<i>Implementasi Hasil Klasifikasi Bidang Dosen</i>	70
BAB V	PENUTUP.....	71
5.1	KESIMPULAN.....	71
5.2	SARAN.....	72
DAFTAR PUSTAKA.....		73

DAFTAR TABEL

Table 2. 1 Studi Literatur	8
Table 2. 2 Tabel <i>Confusion Matrix</i>	16
Table 3. 1 Hardware (perangkat keras)	21
Table 3. 2 Software (perangkat lunak)	22
Table 3. 3 Sample Data Utuh	24
Table 3. 4 Sample Penambahan Dataset Baru	26
Table 3. 5 Sample Data Utuh ditambah Data Bidang Dosen	28
Table 3. 6 Mereduksi Data	30
Table 3. 7 Melabeli atau Mengelompokkan Bidang Dosen	31
Table 3. 8 Proses Case Folding	32
Table 3. 9 Menghilangkan angka	33
Table 3. 10 Menghilangkan tanda baca	35
Table 3. 11 Menghilangkan Spasi yang tidak perlu	36
Table 3. 12 Proses Stopword	37
Table 3. 13 Sample Data yang telah di preprocessing	40
Table 3. 14 Perhitungan TF	41
Table 3. 15 Perhitungan IDF	44
Table 3. 16 Perhitungan TF-IDF	48
Table 3. 17 Contoh Kasus	51
Table 3. 18 Perhitungan Panjang Vector	52
Table 3. 19 Urutan hasil perhitungan <i>cosine similarity</i>	57
Table 3. 20 Hasil Kemiripan Terdekat	57
Table 4. 1 Perbandingan Pengujian	67

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	23
Gambar 3. 2 Pembagian Dataset.....	39
Gambar 3. 3 Rancangan Design Interface.....	58
Gambar 4. 1 Dataset di Excel.....	60
Gambar 4. 2 Dataset di Jupyter Notebook.....	60
Gambar 4. 3 Data 'Judul Skripsi' dan 'Bidang Dosen'.....	61
Gambar 4. 4 Pelabelan Bidang Dosen.....	62
Gambar 4. 5 Source Code Load Data di Jupyter Notebook.....	62
Gambar 4. 6 Source Code Text Preprocessing Case Folding.....	63
Gambar 4. 7 Source Code menghilangkan link dan angka.....	63
Gambar 4. 8 Source Code menghilangkan tanda baca.....	63
Gambar 4. 9 Source Code menghilangkan spasi berlebihan.....	64
Gambar 4. 10 Source Code proses Stopword.....	64
Gambar 4. 11 Source Code Data Split.....	65
Gambar 4. 12 Source Code proses Encoder.....	65
Gambar 4. 13 Source Code proses TF-IDF.....	65
Gambar 4. 14 Source Code proses klasifikasi algoritma K-NN.....	66
Gambar 4. 15 Source Code Confusion Matrix.....	66
Gambar 4. 16 Confusion Matrix di Jupyter Notebook.....	67
Gambar 4. 17 Implementasi Website Rekomendasi Bidang Dosen.....	69
Gambar 4. 18 Implementasi Input Judul Skripsi.....	70
Gambar 4. 19 Hasil Klasifikasi Bidang Dosen.....	70

INTISARI

Pada jenjang pendidikan khususnya untuk S1, syarat kelulusan adalah menyelesaikan skripsi. Dalam menyusun skripsi, mahasiswa didampingi oleh dosen pembimbing yang akan mengarahkan serta sebagai tempat berkonsultasi. Kasusnya masih ada mahasiswa yang bingung untuk mengambil skripsi. Ada beberapa alasan yaitu mereka belum mempunyai judul yang akan diajukan, dan bingung memilih dosen pembimbing yang sesuai dengan judul atau tema mereka. Terkadang didalam kampus, mahasiswa bisa mendapatkan dosen pembimbing, namun tidak sesuai dengan bidangnya, dan tidak sesuai dengan tema judul skripsi. Oleh karena itu, dalam penelitian ini akan membuat klasifikasi bidang dosen berdasarkan judul mahasiswa.

Data yang digunakan sebanyak 1598 diambil dari pihak kampus Universitas AMIKOM Yogyakarta dengan menambahkan data baru sebanyak 15 sehingga jumlah data menjadi 1613. Dengan adanya dosen pembimbing sesuai dengan bidangnya maka akan mempermudah membimbing mahasiswa. Penelitian ini melakukan tahapan labeling, *text preprocessing*, dan pembobotan kata atau disebut dengan TF-IDF (*Term Frequency – Inverse Document Frequency*).

Dataset dibagi menjadi data training dan data test sebanyak 8:2 (80% data training dan 20% data test), lalu data tersebut diklasifikasi dengan algoritma K-NN (*K-Nearest Neighbor*). Penelitian ini mencari jarak tetangga terdekat antar dokumen. Adapun hasil akurasi yang didapat dalam proses klasifikasi yaitu sebesar 85.14% dengan mengambil nilai k sebanyak 18 ($k=18$).

Kata Kunci: Skripsi, Bidang Dosen, K-NN

ABSTRACT

At the level of education, especially for SI, the graduation requirement is to complete the thesis. In preparing the thesis, students are accompanied by a guidance lecturer who will direct and as a place to consult. The case is still there are students who are confused to take a thesis. There are several reasons that they do not have a title to be submitted, and are confused to choose a tutor who matches their title or theme. Sometimes on campus, students can get a guidance lecturer, but not in accordance with the field, and not in accordance with the theme of the thesis title. Therefore, in this study will make a classification of lecturer fields based on student titles.

The data used as many as 1598 was taken from the campus of AMIKOM Yogyakarta University by adding 15 new data so that the amount of data became 1613. With the lecturer in accordance with the field, it will be easier to guide students. This study conducted stages of labeling, text preprocessing, and word weighting or called TF-IDF (Term Frequency – Inverse Document Frequency).

The dataset is divided into training data and test data as much as 8:2 (80% training data and 20% test data), then the data is classified with the K-NN (K-Nearest Neighbor) algorithm. The study looked for the closest neighbors between documents. The results of accuracy obtained in the classification process is 85.14% by taking a value of k as much as 18 ($k = 18$).

Keywords: Undergraduate, Lecturer Field, K-NN