

**TUGAS AKHIR**

**KLASIFIKASI *MULTI-LABEL* TEKS HADIST  
DENGAN MENGGUNAKAN *K-NEAREST NEIGHBOR*  
DAN DIINTEGRASIKAN *LABEL POWERSET***



**IAN APRILIANTO SAID  
16102196**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS INFORMATIKA  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO  
2023**

**TUGAS AKHIR**

**KLASIFIKASI *MULTI-LABEL* TEKS HADIST  
DENGAN MENGGUNAKAN *K-NEAREST NEIGHBOR*  
DAN DIINTEGRASIKAN *LABEL POWERSET***

***MULTI-LABEL* CLASSIFICATION OF HADITH  
TEXTS USING *K-NEAREST NEIGHBOR* AND  
INTEGRATED WITH *POWERSET* LABELS**

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



**IAN APRILIANTO SAID  
16102196**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS INFORMATIKA  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO  
2023**

**LEMBAR PERSETUJUAN PEMBIMBING**

**KLASIFIKASI *MULTI-LABEL* TEKS HADIST  
DENGAN MENGGUNAKAN K-NEAREST NEIGHBOR  
DAN DIINTEGRASIKAN LABEL POWERSET**

***MULTI-LABEL CLASSIFICATION OF HADITH  
TEXTS USING K-NEAREST NEIGHBOR AND  
INTEGRATED WITH POWERSET LABELS***

Dipersiapkan dan Disusun Oleh

IAN APRILIANTO SAID  
16102196

**Fakultas Informatika  
Institut Teknologi Telkom Purwokerto  
Pada Tanggal : 01 Agustus 2023**

**Pembimbing Utama,**



**Rifki Adhitama, S.Kom., M.Kom  
NIDN. 0627089101**

**KLASIFIKASI *MULTI-LABEL* TEKS HADIST  
DENGAN MENGGUNAKAN K-NEAREST NEIGHBOR  
DAN DIINTEGRASIKAN LABEL POWERSSET**

***MULTI-LABEL* CLASSIFICATION OF HADITH  
TEXTS USING K-NEAREST NEIGHBOR AND  
INTEGRATED WITH POWERSSET LABELS**

Disusun Oleh  
IAN APRILIAN TO SAID  
16102196

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas  
Akhir Pada Rabu, Tanggal 16 Agustus 2023.

**Penguji I**



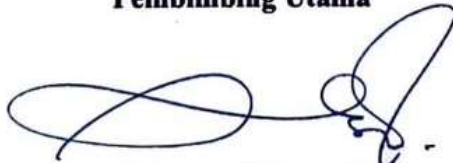
**Auliya Burhanuddin, S.Si., M.Kom.**  
NIDN. 0630058202

**Penguji II**



**Dasril Alde, S.Kom., M.Kom.**  
NIDN. 1026049401

**Pembimbing Utama**



**Rifki Adhitama, S.Kom., M.Kom.**  
NIDN. 0627089101

**Dekan,**



**Auliya Burhanuddin, S.Si., M.Kom**  
NIK. 19820008

## HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

**Nama Mahasiswa** : Ian Aprilianto Said  
**NIM** : 16102196  
**Program Studi** : Teknik Informatika

Menyatakan bahwa Tugas Akhir dengan judul berikut:

**KLASIFIKASI *MULTI-LABEL* TEKS HADIST DENGAN  
MENGUNAKAN *K-NEAREST NEIGHBOR* DAN DIINTEGRASIKAN  
*LABEL POWERSET***

Dosen Pembimbing Utama : Rifki Adhitama, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

**Purwokerto, 01 Agustus 2023,**

**Yang Menyatakan,**



## KATA PENGANTAR

Tidak ada kata lain selain mengucapkan puji syukur atas terselesaikannya skripsi ini dengan judul "**Klasifikasi *Multi-label* Teks Hadist Dengan Menggunakan *K-Nearest Neighbor* Dan Diintegrasikan *Label PowerSet***".

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kata sempurna karena keterbatasan kemampuan dan pengetahuan yang penulis dapatkan oleh karena itu dengan kerendahan hati penulis mohon maaf atas segala kekurangan.

Penyusunan skripsi ini tidak akan berhasil tanpa ada bantuan dan kerja sama dari pihak lain. Oleh karenanya kesempatan ini penulis ingin mengucapkan terima kasih sebesar-besarnya kepada semua pihak yang telah membantu dan mendorong terwujudnya skripsi ini. Segala kerendahan hati penulis mengucapkan terima kasih khususnya kepada:

1. Ibu Dr. Tenia Wahyuningrum, S.Kom., M.T selaku Rektor Institut Teknologi Telkom Purwokerto.
2. Bapak Auliya Burhanuddin, S.Si., M.Kom selaku Dekan Fakultas Informatika.
3. Ibu Amalia Beladinna Arifa, S.Pd, M.Cs selaku Kepala Program Studi Teknik Informatika Institut Teknologi Telkom Purwokerto.
4. Bapak Rifki Adhitama, S.Kom., M. yang penuh kesabaran membimbing dan mengarahkan penulis sehingga dapat terselesaikannya skripsi ini.
5. Kedua orang tua penulis yang selalu memberikan motivasi dan juga rekan-rekan penulis yang tidak henti-hentinya memberikan dukungan motivasi, semangat dan dukungan mental.

Purwokerto, 01 Agustus 2023



Ian Aprilianto Said

## DAFTAR ISI

TUGAS AKHIR.....	i
TUGAS AKHIR.....	ii
LEMBAR PERSETUJUAN PEMBIMBING.....	iii
LEMBAR PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	x
ABSTRAK.....	xi
ABSTRACT.....	xii
BAB I PENDAHULUAN.....	13
<b>1.1 Latar Belakang Masalah</b> .....	13
<b>1.2 Rumusan Masalah</b> .....	16
<b>1.3 Tujuan Penelitian</b> .....	16
<b>1.4 Batasan Penelitian</b> .....	16
<b>1.5 Manfaat Penelitian</b> .....	17
BAB II TINJAUAN PUSTAKA.....	18
<b>2.1 Penelitian Sebelumnya</b> .....	18
<b>2.2 Dasar Teori</b> .....	22
<b>2.2.1 Multi-label Classification</b> .....	22
<b>2.2.2 K-Nearest Neighbor</b> .....	22
<b>2.2.3 Label Powerset</b> .....	24
<b>2.2.4 TF-IDF</b> .....	24
<b>2.2.5 Preprocessing</b> .....	26
<b>2.2.6 K-fold Cross Validation</b> .....	27
<b>2.2.7 Hamming loss</b> .....	28
<b>2.2.8 Python</b> .....	29
<b>2.2.9 Natural Language Processing</b> .....	29
<b>2.2.10 Binary Relevance</b> .....	29

BAB III METODOLOGI PENELITIAN.....	31
<b>3.1 Objek dan Subjek Penelitian</b> .....	31
<b>3.2 Alat dan Bahan Penelitian</b> .....	31
3.2.1 Alat penelitian .....	31
3.2.2 Bahan penelitian.....	31
<b>3.3 Diagram Alir Penelitian</b> .....	32
3.3.1. Studi Literatur.....	32
3.3.2. Pengumpulan Data.....	33
3.3.3. <i>Preprocessing</i> .....	33
3.3.4. Ekstraksi Fitur .....	35
3.3.5. <i>Problem Transformation</i> .....	36
3.3.6. Pembagian Data .....	37
3.3.7. Klasifikasi .....	37
3.3.8. Evaluasi .....	38
BAB IV HASIL PENGUJIAN DAN ANALISIS.....	39
<b>4.1 Hasil Pengujian</b> .....	39
4.1.1 Hasil Pengumpulan Data .....	39
4.1.2 Hasil <i>Preprocessing</i> Data .....	39
4.1.3 Hasil Ekstraksi Fitur .....	41
4.1.4 Hasil <i>Transformation Problem</i> .....	42
4.1.5 Hasil Pemisahan Data .....	43
4.1.6 Hasil Klasifikasi.....	43
4.1.7 Pengujian .....	46
<b>4.2 Pembahasan dan Analisa</b> .....	50
BAB V KESIMPULAN DAN SARAN.....	58
<b>5.1 Kesimpulan</b> .....	58
<b>5.2 Saran</b> .....	58
DAFTAR PUSTAKA .....	60
LAMPIRAN.....	62



## DAFTAR GAMBAR

Gambar 2. 1	Gambaran klasifikasi k-nearest neighbor .....	23
Gambar 2. 2	Gambaran ransformasi label powerset .....	24
Gambar 2. 3	Gambaran proses k-fold cross validation .....	28
Gambar 3. 1	Flowchart alur penelitian.....	32
Gambar 4.1	Grafik hamming loss pengujian 1 .....	48
Gambar 4.2	Hasil grafik label powerset.....	49
Gambar 4.3	Hasil grafik binary relevance .....	49
Gambar 4.4	Grafik Data Label 1 .....	56
Gambar 4.5	Gambar Grafik Label 2 .....	57

## DAFTAR TABEL

Tabel 1.1 <i>Dataset</i> dari penelitian sebelumnya .....	14
Tabel 2.1 Penelitian Sebelumnya.....	20
Tabel 3. 1 Tahapan case folding .....	33
Tabel 3. 2 Tahapan stopwords removal.....	34
Tabel 3. 3 Tahapan tokenization .....	34
Tabel 3. 4 Tahapan stemming .....	34
Tabel 3. 5 Term frequency .....	35
Tabel 3. 6 Inverse document frequency .....	36
Tabel 3. 7 Term frequency dan inverse document frequency .....	36
Tabel 3. 8 <i>Dataset</i> sebelum label powerset.....	37
Tabel 3. 9 <i>Dataset</i> setelah label powerset.....	37
Tabel 3. 10 Tahapan k-fold cross validation.....	37
Tabel 4. 1 <i>Dataset</i> .....	39
Tabel 4. 2 <i>Dataset</i> proses case folding.....	39
Tabel 4. 3 <i>Dataset</i> proses stopwords removal .....	40
Tabel 4. 4 <i>Dataset</i> proses tokenization .....	40
Tabel 4. 5 <i>Dataset</i> proses stemming .....	41
Tabel 4. 6 <i>Dataset</i> tf-idf.....	41
Tabel 4. 7 <i>Dataset multi-label</i> .....	42
Tabel 4. 8 <i>Dataset</i> multi-class.....	42
Tabel 4. 9 Data tabel train knn .....	44
Tabel 4. 10 Data tabel testing knn.....	44
Tabel 4. 11 Data urutan euclidean distance .....	45
Tabel 4. 12 Data voting euclidean distance .....	45
Tabel 4. 13 Data hasil prediksi knn.....	46
Tabel 4. 14 Hasil pengujian skenario 1 .....	47
Tabel 4. 15 Hasil pengujian skenario 2.....	48