

**KLASIFIKASI MULTI-LABEL PADA ABSTRAK SKRIPSI
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES**

**(Studi Kasus Program Studi Informatika Universitas Amikom
Yogyakarta)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AJI SYAHRONI ARDIANTO

16.11.0508

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

**KLASIFIKASI MULTI-LABEL PADA ABSTRAK SKRIPSI
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES**

**(Studi Kasus Program Studi Informatika Universitas Amikom
Yogyakarta)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AJI SYAHRONI ARDIANTO

16.11.0508

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI MULTI-LABEL PADA ABSTRAK SKRIPSI
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES**

(Studi Kasus Program Studi Informatika Universitas Amikom Yogyakarta)

yang disusun dan diajukan oleh

Aji Syahroni Ardianto

16.11.0508

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 November 2022

Dosen Pembimbing.



Sumarni Adi, S.Kom, M.Cs

NIK. 190302256

HALAMAN PENGESAHAN

SKRIPSI

KLASIFIKASI MULTI-LABEL PADA ABSTRAK SKRIPSI
MENGUNAKAN METODE MULTINOMIAL NAIVE BAYES

(Studi Kasus Program Studi Informatika Universitas Amikom Yogyakarta)

yang disusun dan diajukan oleh

Aji Syahroni Ardianto

16.11.0508

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 November 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Donni Prubowo, M.Kom
NIK. 190302253

Yoga Pristyanto, S.Kom, M.Eng
NIK. 190302412

Arif Akbarul Huda, S.Si, M.Eng
NIK. 190302287



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 November 2022

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aji Syahroni Aridanto

NIM : 16.11.0508

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Sumarni Adi, S.Kom, M.Cs

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari 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 tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian 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.

Yogyakarta, 23 November 2022

Yang Menyatakan,



The image shows a handwritten signature in black ink over a purple and yellow circular stamp. The stamp contains the text 'METERAI TEMPEL' and a unique identification number 'IC4AKX391380777'.

Aji Syahroni Ardianto

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan pada Tuhan Yang Maha Esa yang telah melimpahkan karunia-Nya sehingga penulis dapat menyusun dan menyelesaikan Skripsi. Penulis pun tak lupa mengucapkan serangkaian ucapan terimakasih kepada seluruh pihak yang telah membantu penulis, baik yang terlibat secara langsung maupun tidak langsung. Skripsi ini penulis persembahkan kepada:

1. Keluarga besar yang senantiasa memberikan semangat, doa, harapan, nasihat dan dukungan.
2. Ibu Sumarni Adi, S.Kom, M.Cs selaku Dosen Pembimbing yang telah bersedia memberikan waktu, saran, ilmu, dan bimbingan.
3. Semua teman-teman yang selalu memberikan semangat dan motivasi yang tidak dapat penulis sebutkan satu persatu.



KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan yang telah melimpahkan kebaikan dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**Klasifikasi Multi-Label pada Abstrak Skripsi Menggunakan Metode Multinomial Naive Bayes**". Skripsi ini disusun untuk memenuhi salah satu persyaratan dalam menyelesaikan Program Studi Strata-1 Informatika di Universitas Amikom Yogyakarta. Rasa terima kasih penulis sampaikan kepada seluruh pihak yang telah membantu, membimbing, dan mendukung, khususnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Sumarni Adi, S.Kom, M.Cs selaku Dosen Pembimbing.
3. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta.
4. Seluruh teman-teman Mahasiswa/i Universitas Amikom Yogyakarta.

Penulis menyadari bahwa masih terdapat banyak kekurangan di dalam skripsi ini. Untuk itu kritik dan saran akan diterima dengan senang hati. Semoga adanya skripsi ini dapat memberi manfaat dan wawasan kepada pembaca.

Yogyakarta, 19 Juni 2023

Penulis

DAFTAR ISI

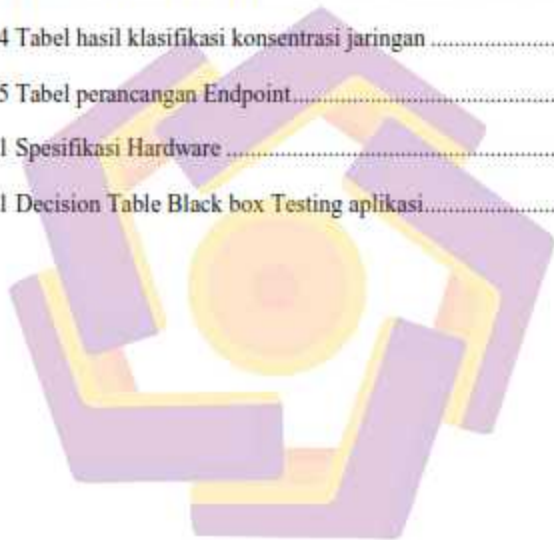
HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN.....	xiv
DAFTAR ISTILAH.....	xvi
INTISARI.....	xix
ABSTRACT.....	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Studi Literatur.....	9

2.2	Dasar Teori.....	13
2.2.1	Text Mining.....	13
2.2.2	Klasifikasi.....	15
2.2.3	Klasifikasi Multi-Label	16
2.2.4	Multinomial Naive Bayes.....	19
2.2.5	Pengujian.....	19
2.2.6	Program Studi Informatika.....	20
2.2.7	Abstrak Skripsi.....	20
2.2.8	Topik Penelitian Skripsi Informatika	21
2.2.9	Fast API.....	28
2.2.10	Database SQLite	28
2.2.11	Antarmuka	28
BAB III METODE PENELITIAN.....		29
3.1	Objek Penelitian.....	29
3.2	Tinjauan Umum	29
3.3	Alur Penelitian	30
3.3.1	Data acquisition.....	31
3.3.2	Labeling.....	32
3.3.3	Preprocessing.....	32
3.3.4	Pembagian Dataset	34
3.3.5	Transformasi Data / Pembobotan Kata	34
3.3.6	Klasifikasi.....	34
3.3.7	Evaluasi Hasil Klasifikasi	34
3.3.8	Implementasi Model Klasifikasi	35
3.3.9	Perancangan Aplikasi	35

3.4	Alat dan Bahan Penelitian.....	42
3.4.1	Hardware	42
3.4.2	Software.....	43
BAB IV HASIL DAN PEMBAHASAN		44
4.1	Data Acquisition.....	44
4.2	Data Labeling.....	47
4.3	Data Preprocessing.....	50
4.3.1	Cleansing	50
4.3.2	Filtering	52
4.3.3	Stemming.....	53
4.3.4	Tokenizing.....	54
4.4	Pembobotan Kata	55
4.5	Pembagian Dataset.....	57
4.6	Klasifikasi	59
4.7	Evaluasi Klasifikasi.....	60
4.8	Implementasi Model Klasifikasi	62
4.9	Perancangan Aplikasi.....	67
4.9.1	Perancangan Antarmuka.....	67
4.9.2	Pembuatan REST API.....	73
4.9.3	Implementasi REST API.....	79
4.10	Pengujian Aplikasi	82
BAB V PENUTUP.....		85
5.1	Kesimpulan	85
5.2	Saran.....	85
REFERENSI		87

DAFTAR TABEL

Tabel 2. 1 Perbandingan penelitian terdahulu.....	11
Tabel 3. 2 Tabel hasil klasifikasi konsentrasi pemrograman	38
Tabel 3. 3 Tabel hasil klasifikasi konsentrasi multimedia	39
Tabel 3. 4 Tabel hasil klasifikasi konsentrasi jaringan	39
Tabel 3. 5 Tabel perancangan Endpoint.....	40
Tabel 3. 1 Spesifikasi Hardware	43
Tabel 4. 1 Decision Table Black box Testing aplikasi.....	83

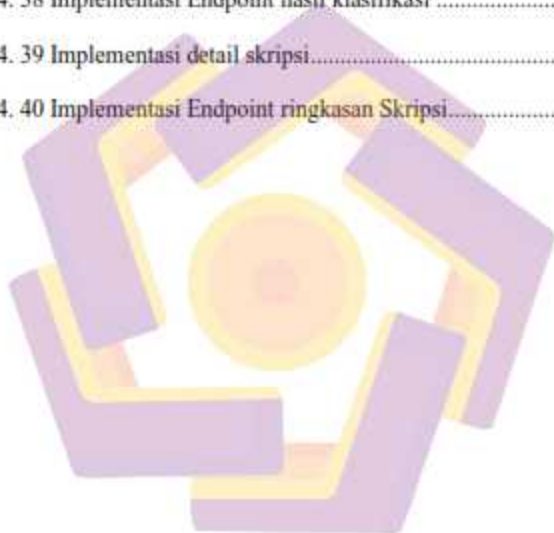


DAFTAR GAMBAR

Gambar 1. 1 Statistik jumlah tugas akhir dan skripsi	1
Gambar 1. 2 Alur metodologi penelitian.....	5
Gambar 2. 1 Sampel dataset Multi-Label Binary Relevance	16
Gambar 2. 2 Sampel proses Binary Relevance	17
Gambar 2. 3 Sampel dataset Multi-Label Classifier Chain.....	17
Gambar 2. 4 Sampel proses Classifier Chain.....	18
Gambar 2. 5 Sampel dataset Multi-Label Label Powerset.....	18
Gambar 2. 6 Sampel proses Label Powerset.....	18
Gambar 2. 7 Topik utama pada CSO	22
Gambar 3. 1 Alur metodologi penelitian.....	31
Gambar 3. 2 Desain halaman landing	36
Gambar 3. 3 Desain halaman hasil klasifikasi	37
Gambar 3. 4 Desain halaman detail penelitian.....	37
Gambar 3. 5 Desain halaman ringkasan.....	38
Gambar 4. 1 Penyajian data pada website Eprints.....	44
Gambar 4. 2 Sample JSON	45
Gambar 4. 3 Source Code penyesuaian kolom dan penyatuan data	46
Gambar 4. 4 Source Code penghapusan data redundan dan penyeleksian data ...	47
Gambar 4. 5 Sampel hasil transformasi dataset menjadi Multi-Label	47
Gambar 4. 6 Distribusi topik penelitian	48
Gambar 4. 7 Korelasi pearson topik penelitian pada data terlabel.....	50
Gambar 4. 8 Sekumpulan prosedur cleansing.....	51

Gambar 4. 9 Implementasi prosedur cleansing	52
Gambar 4. 10 Sekumpulan prosedur filtering dan pengimplementasiannya.	53
Gambar 4. 11 Prosedur Stemming dan pengimplementasiannya.....	54
Gambar 4. 12 Prosedur Tokenizing	55
Gambar 4. 13 Sampel pengurangan kata dari hasil preprocessing.....	55
Gambar 4. 14 Pengimplementasian TF-IDF dengan SKLEARN	56
Gambar 4. 15 Grafik frekuensi kemunculan kata	57
Gambar 4. 16 Sampel perhitungan TF-IDF	57
Gambar 4. 17 Implementasi pembagian dataset untuk ketiga konsentrasi	58
Gambar 4. 18 Distribusi jumlah skripsi per topik	59
Gambar 4. 19 Implementasi klasifikasi.....	60
Gambar 4. 20 Implementasi perhitungan Hamming Loss	61
Gambar 4. 21 Hasil perbandingan Hamming Loss ketiga metode.....	62
Gambar 4. 22 Implementasi model dan penyesuaian kolom	65
Gambar 4. 23 Implementasi penambahan kolom subfield.....	66
Gambar 4. 24 Implementasi transformasi data frame ke dalam bentuk database .	67
Gambar 4. 25 Sampel tabel	67
Gambar 4. 26 Highlight Source Code antarmuka landing page.....	68
Gambar 4. 27 Hasil tampilan landing page	69
Gambar 4. 28 Highlight Source Code halaman klasifikasi	70
Gambar 4. 29 Tampilan halaman klasifikasi.....	71
Gambar 4. 30 Highlight Source Code halaman ringkasan	72
Gambar 4. 31 Tampilan halaman ringkasan	73

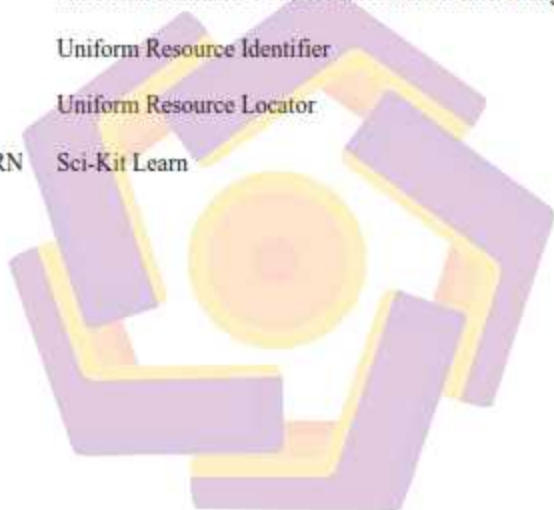
Gambar 4. 32 Logic Endpoint hasil klasifikasi	74
Gambar 4. 33 Response JSON hasil klasifikasi	75
Gambar 4. 34 Logic Endpoint detail skripsi	76
Gambar 4. 35 Response JSON detail skripsi	77
Gambar 4. 36 Logic ringkasan topik skripsi	78
Gambar 4. 37 Response JSON ringkasan topik skripsi.....	79
Gambar 4. 38 Implementasi Endpoint hasil klasifikasi	80
Gambar 4. 39 Implementasi detail skripsi.....	81
Gambar 4. 40 Implementasi Endpoint ringkasan Skripsi.....	82



DAFTAR LAMBANG DAN SINGKATAN

\propto	Sebanding dengan
Π	Huruf besar Pi
\leq	Kurang dari atau sama dengan
$<$	Kurang dari
\neq	Tidak sama dengan
Σ	Sigma atau penjumlahan
SVM	Support Vector Machine
NBC	Naïve Bayes Classifier
KNN	K-Nearest Neighbor
CSO	Computer Science Ontology
ACM	Association for Computing Machinery
API	Application Programming Interface
JSON	Javascript Object Notation
TF-IDF	Term Frequency-Inverse Document Frequency
SDLC	System Development Life Cycle
PT	Problem Transformation
AA	Algorithm Adaptation
AI	Artificial Intelligence
2D	2 Dimensi
3D	3 Dimensi
IT	Information and Technology

NLTK	Natural Language Tool Kit
UI/UX	User Interface / User Experience
REST API	Representational State Transfer Application Programming Interface
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
CSV	Comma Separated Values
ASCII	American Standard Code for Information Interchange
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
SKLEARN	Sci-Kit Learn



DAFTAR ISTILAH

Artificial Intelligence	kecerdasan buatan
Big Data Analytics	analisis pada data yang berukuran besar
Data Science	sains data
Data Mining	pencarian informasi pada data
Text classification	klasifikasi data teks
Supervised Learning	pembelajaran terarah
Classifier	pengklasifikasi
Data Acquisition	akuisisi data dari suatu sumber
Preprocessing	tahapan praproses
Feature	fitur
Feature Selection	pemilihan fitur
Feature Extraction	ekstraksi fitur
Classification Stage	tahapan Pengklasifikasian
Data Splitting	tahapan pembagian dataset
Data Training	data yang digunakan dalam proses pelatihan
Data Testing	data yang digunakan dalam proses pengujian
Data Frame	data tabular 2 dimensi
Imbalance Data	ketidakseimbangan kelas target pada data
Recall	rasio prediksi data positif yang benar dengan keseluruhan hasil yang diprediksi positif.

Accuracy	rasio prediksi benar negatif dan positif dengan semua data
Precision	rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif
Framework	kerangka kerja yang digunakan untuk mengembangkan sistem
Interface	perangkat atau sistem yang digunakan untuk mengakses program
Frontend	komposisi tampilan dari sebuah sistem
Backend	komposisi fungsionalitas yang berkaitan langsung dengan database
Functional Testing	pengujian fungsionalitas sistem
Review	ulasan
Output	hasil dari sebuah proses atau tahapan
Noise	gangguan yang memicu ketidakakuratan pada hasil
Term	kata atau token pada sebuah kalimat pada dataset
Robust	sebuah sistem yang kuat
Syntax	perintah pada program
Full Featured	fitur lengkap yang dimiliki sistem atau program
Open Source	sistem atau program yang dipublikasikan secara umum
Self-contained	sistem atau program yang bersifat mandiri
Cross Platform	sistem yang dapat diakses di berbagai platform
Device	perangkat yang digunakan untuk menjalankan sistem

User Experience	pengalaman pengguna saat menggunakan aplikasi atau sistem
Decision Tables	tabel yang digunakan untuk penunjang keputusan
Library	pustaka kode yang digunakan untuk mempermudah pengembangan sistem
Scope	cakupan dari sebuah sistem
Endpoint	titik akhir/prosedur API yang diakses dari sistem eksternal
Test Case	kasus yang digunakan untuk pengujian sistem
Tools	sekumpulan pustaka kode untuk mempermudah pengembangan
Labeling	tahapan yang dilakukan untuk memberi label pada sebuah data
Filtering	tahapan yang dilakukan untuk mengeliminasi entitas yang tidak diperlukan
Landing Page	halaman pada website yang memiliki tujuan kampanye
Highlight source code	cuplikan dari sekumpulan kode yang menjalankan fungsionalitas tertentu
Keyword	kata kunci
Logic	kumpulan kode perintah yang digunakan untuk menyelesaikan permasalahan
Mapping JSON	proses pemetaan/penguraian pada tipe data JSON
Parsing JSON	proses perubahan pada tipe data JSON ke dalam tipe data tertentu

INTISARI

Program studi informatika merupakan program studi yang berada pada lingkungan fakultas informatika Universitas Amikom Yogyakarta. Salah satu misi dari Informatika adalah peningkatan dan perbaikan pada bidang penelitian. Salah satu bidang penelitian adalah skripsi, data skripsi mengalami penambahan setiap tahunnya. Topik skripsi yang diambil mahasiswa umumnya sesuai dengan mata kuliah bidang konsentrasi, akan tetapi mata kuliah bidang konsentrasi selalu mengikuti perkembangan teknologi. Hal ini tentunya berpengaruh terhadap topik skripsi yang diambil oleh mahasiswa. Semakin bervariasinya topik yang diambil mahasiswa membuat pengelompokan skripsi berdasarkan topik relatif sulit. Klasifikasi merupakan salah satu metode yang dapat digunakan untuk memperoleh informasi baru dari kasus ini. Klasifikasi yang sesuai dengan kasus ini adalah klasifikasi multi label. Metode klasifikasi multi label yang dipilih dalam penelitian ini adalah *Problem Transformation* dengan pendekatan *Binary Relevance*, *Classifier Chain*, dan *Label Powerset* dengan *Multinomial Naive Bayes* sebagai estimatornya. Dari ketiga metode *Problem Transformation* dipilih satu metode terbaik untuk membangun aplikasi berbasis web. Evaluasi klasifikasi multi label dilakukan dengan pengukuran nilai akurasi dan Hamming Loss. Hasil dari penelitian ini adalah Algoritma Multinomial Naive Bayes yang dikombinasikan dengan Classifier Chain cenderung memberikan kinerja terbaik dengan akurasi diatas 85% dan Hamming Loss dibawah 0.037.

Kata kunci: Klasifikasi Multi-Label, Problem Transformation, Hamming Loss.

ABSTRACT

The informatics study program is a study program within the informatics faculty of Amikom Yogyakarta University. One of the missions of Informatics is improvement and improvement in the field of research. One of the areas of research is thesis, thesis data has increased every year. Thesis topics taken by students are generally in accordance with the concentration field courses, however, the concentration field courses always follow technological developments. This certainly affects the thesis topics taken by students. The increasing variety of topics taken by students makes grouping theses based on topics relatively difficult. Classification is one method that can be used to obtain new information from this case. The classification that fits this case is the multi-label classification. The multi-label classification method chosen in this study is Problem Transformation with the Binary Relevance, Classifier Chain, and Label Powerset approaches with Multinomial Naive Bayes as the estimator. Of the three Problem Transformation methods, the best method was chosen to build web-based applications. Evaluation of multi-label classification is carried out by measuring the value of accuracy and Hamming Loss. The result of this research is that the Naive Bayes Multinomial Algorithm combined with Classifier Chain tends to give the best performance with accuracy above 85% and Hamming Loss below 0.037.

Keyword: Multi-Label Classification, Problem Transformation, Hamming Loss