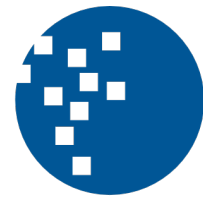


**KLASIFIKASI LIRIK LAGU BERDASARKAN KONTEN
SEKSUAL MENGGUNAKAN VOTING CLASSIFIER**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

LIUSIANTO
00000061560

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

**KLASIFIKASI LIRIK LAGU BERDASARKAN KONTEN
SEKSUAL MENGGUNAKAN VOTING CLASSIFIER**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**LIUSIANTO
00000061560**

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Liusianto
Nomor Induk Mahasiswa : 00000061560
Program Studi : Informatika

Skripsi dengan judul:

Klasifikasi Lirik Lagu Berdasarkan Konten Seksual Menggunakan Voting Classifier

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, Kamis, 03 Juli 2025



(Liusianto)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

KLASIFIKASI LIRIK LAGU BERDASARKAN KONTEN SEKSUAL MENGUNAKAN VOTING CLASSIFIER

oleh

Nama : Liusianto
NIM : 00000061560
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 17 Juli 2025

Pukul 10.00 s/d 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

(Dr. Maria Irmira Prasetyowati, S.Kom.,
M.T.)

NIDN: 0725057201

Penguji

(Januar Wahjudi, S.Kom., M.Sc.)

NIDN: 0330017201

Pembimbing

(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Liusianto
NIM : 00000061560
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Klasifikasi Lirik Lagu Berdasarkan
Konten Seksual Menggunakan Voting
Classifier

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, Kamis, 03 Juli 2025

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Yang menyatakan



Liusianto

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Klasifikasi Lirik Lagu Berdasarkan Konten Seksual Menggunakan Voting Classifier dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai penyusunan laporan skripsi ini, sangat sulit bagi saya untuk menyelesaikan laporan skripsi ini. Oleh sebab itu, saya ingin mengucapkan terima kasih kepada:

Mengucapkan terima kasih

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat menjadi informasi yang berguna bagi

Tangerang, Kamis, 03 Juli 2025

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Liusianto

KLASIFIKASI LIRIK LAGU BERDASARKAN KONTEN SEKSUAL MENGUNAKAN VOTING CLASSIFIER

Liusianto

ABSTRAK

Industri televisi dan radio di Indonesia merupakan bagian penting dari industri Indonesia sebagai penyebar informasi. Persaingan yang tinggi menyebabkan banyak siaran yang mengabaikan nilai-nilai yang dianut masyarakat. Salah satu bagian dari siaran adalah siaran lagu. Siaran lagu yang mengabaikan kualitas dapat menyebabkan lagu dengan konten seksual disiarkan. Untuk mengatasi permasalahan ini, penelitian ini dibuat dengan tujuan mengembangkan sistem klasifikasi lirik lagu berdasarkan konten seksual. Model yang digunakan merupakan soft voting yang merupakan gabungan dari model ETC dan LSTM. Fitur ekstraksi yang digunakan dalam penelitian menggunakan teknik *TF-IDF*, *Word2Vec*, dan *Doc2Vec*. Hasil dari *soft voting* terbaik adalah akurasi sebesar 98,5%, *precision* 88,0%, *recall* 96,6%, dan *f1-score* 92,2%. Dari hasil ini dapat disimpulkan bahwa *soft voting* dapat mengklasifikasi lirik lagu dengan baik.

Kata kunci: (ETC, Lirik lagu, LSTM, Soft Voting)



**CLASSIFICATION OF SONG LYRICS BASED ON SEXUAL CONTENT
USING VOTING CLASSIFIER**

Liusianto

ABSTRACT

The television and radio industry in Indonesia is an important part of the Indonesian industry as a disseminator of information. High competition causes many broadcasts to ignore the values adopted by society. One part of the broadcast is song broadcasts. Song broadcasts that ignore quality can cause songs with sexual content to be broadcast. To overcome this problem, this study was conducted with the aim of developing a song lyrics classification system based on sexual content. The model used is soft voting which is a combination of the ETC and LSTM models. The extraction features used in the study use the TF-IDF, Word2Vec, and Doc2Vec techniques. The results of the best soft voting are an accuracy of 98.5%, precision 88.0%, recall 96.6%, and f1-score 92.2%. From these results it can be concluded that soft voting can classify song lyrics well.

Keywords: (ETC, LSTM, Soft Voting, Song Lyrics) (in alphabetical order)



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Penelitian Pendahulu	6
2.2 Lirik Lagu	7
2.3 Kata Seksual	7
2.4 Preprocessing	7
2.5 Feature Extraction	8
2.6 K-Fold Cross Validation	9
2.7 Voting Classifier	9
2.8 LSTM	10
2.9 Extra Trees Classifier	10
2.10 Metrik Evaluasi	10
BAB 3 METODOLOGI PENELITIAN	12
3.1 Alur Penelitian	12
3.2 Metode Penelitian	12
3.2.1 Telaah Literatur	12
3.2.2 Import Libraries	13
3.2.3 Data Acquisition	13
3.2.4 Data Preprocessing	14
3.2.5 Feature Extraction	15
3.2.6 Algoritma Klasifikasi	15
3.2.7 Metode Ensemble Learning	16
3.2.8 Testing Model	17
3.2.9 Hasil Prediksi	17
BAB 4 HASIL DAN DISKUSI	18
4.1 Load Dataset	18
4.2 Preprocessing	19
4.2.1 Data Cleaning	19

4.2.2	Tokenizing	19
4.2.3	Stopword Removal	20
4.3	Feature Extraction	22
4.3.1	TF-IDF (Term Frequency - Inverse Document Frequency)	22
4.3.2	Word2Vec	22
4.3.3	Doc2Vec	23
4.4	Modeling	23
4.4.1	Split Data	23
4.4.2	Extra Trees Classifier (ETC)	24
4.4.3	Long Short-Term Memory (LSTM)	24
4.4.4	Soft Voting	25
4.5	Evaluasi Model	26
4.5.1	ETC + TF-IDF	27
4.5.2	ETC + Doc2Vec	28
4.5.3	LSTM + Word2Vec	29
4.5.4	Soft Voting: ETC TF-IDF + LSTM	30
4.5.5	Soft Voting: ETC Doc2Vec + LSTM	31
4.6	Testing Model	31
4.7	Hasil dan Pembahasan	33
BAB 5	SIMPULAN DAN SARAN	36
5.1	Simpulan	36
5.2	Saran	37
DAFTAR PUSTAKA		38



DAFTAR TABEL

Tabel 2.1	Komparasi Penelitian Terkait Deteksi Lirik Lagu Eksplisit	6
Tabel 4.1	Evaluasi Kinerja Model Individu	33
Tabel 4.2	<i>K-Fold Cross Validation Soft Voting</i>	34



DAFTAR GAMBAR

Gambar 3.1	Flowchart Alur Penelitian	12
Gambar 3.2	Import Dataset	13
Gambar 3.3	Preprocessing Database	14
Gambar 3.4	Metode Ensemble Learning	16
Gambar 4.1	<i>Import Dataset</i>	18
Gambar 4.2	<i>Hasil Data Cleaning</i>	19
Gambar 4.3	<i>Hasil Tokenizing</i>	20
Gambar 4.4	<i>Hasil Stopword Removal</i>	21
Gambar 4.5	<i>Hasil Stemming</i>	21
Gambar 4.6	<i>Confusion Matrix ETC + TF-IDF</i>	27
Gambar 4.7	<i>Confusion Matrix ETC + Doc2Vec</i>	28
Gambar 4.8	<i>Confusion Matrix LSTM + Word2Vec</i>	29
Gambar 4.9	<i>Confusion Matrix Soft Voting ETC TF-IDF + LSTM Word2Vec</i>	30
Gambar 4.10	<i>Confusion Matrix Soft Voting ETC Doc2Vec + LSTM Word2Vec</i>	31
Gambar 4.11	<i>Prediksi Model Soft Voting</i>	33
Gambar 4.12	<i>Kurva Fold dalam Soft Voting</i>	34



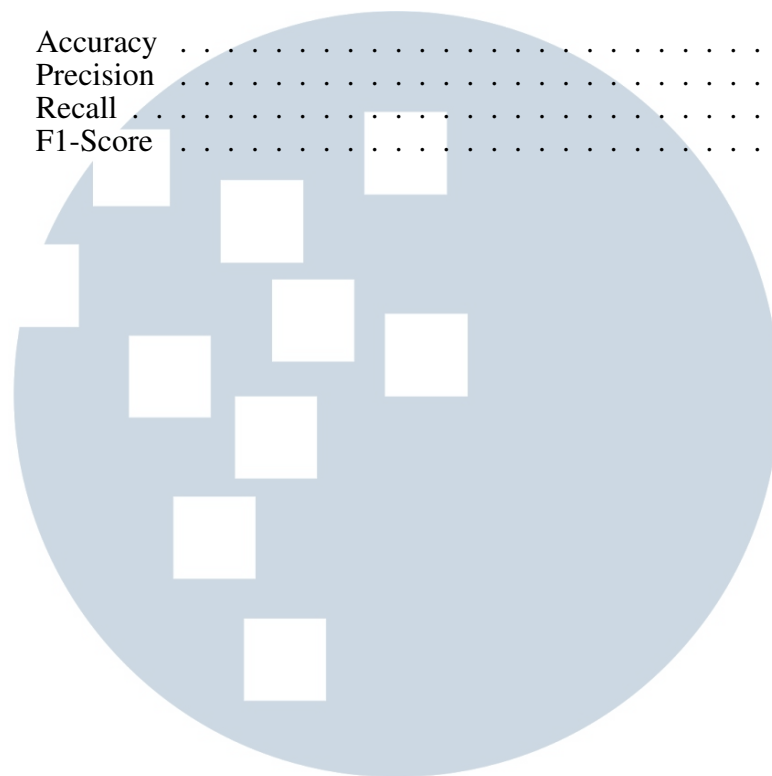
DAFTAR KODE

Kode 4.1	<i>Load Dataset</i>	18
Kode 4.2	<i>Data Cleaning</i>	19
Kode 4.3	<i>Tokenizing</i>	20
Kode 4.4	<i>Stopword Removal</i>	20
Kode 4.5	<i>Stemming</i>	21
Kode 4.6	Ekstraksi fitur <i>TF-IDF</i>	22
Kode 4.7	Pelatihan <i>Word2Vec</i>	22
Kode 4.8	Pelatihan <i>Doc2Vec</i>	23
Kode 4.9	Split Data	23
Kode 4.10	Model <i>Extra Trees Classification</i> dengan <i>TF-IDF</i>	24
Kode 4.11	Model <i>Extra Trees Classification</i> dengan <i>Doc2Vec</i>	24
Kode 4.12	Model LSTM dengan <i>Word2Vec</i>	25
Kode 4.13	<i>Soft Voting ETC TF-IDF dan LSTM Word2Vec</i>	25
Kode 4.14	<i>Soft Voting ETC Doc2Vec dan LSTM Word2Vec</i>	25
Kode 4.15	Prediksi Model <i>Soft Voting</i>	32



DAFTAR RUMUS

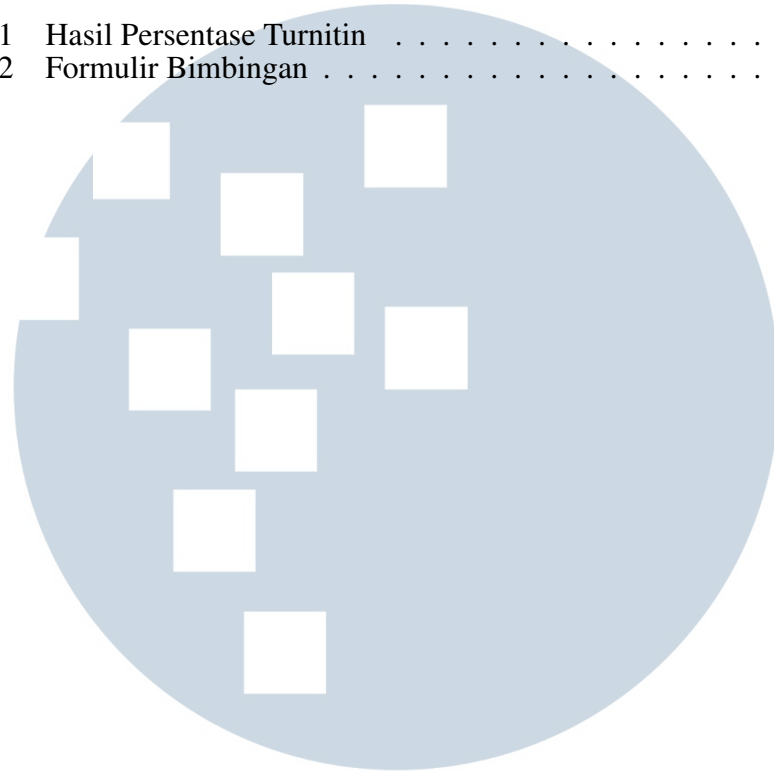
Rumus 2.1	Accuracy	10
Rumus 2.2	Precision	10
Rumus 2.3	Recall	11
Rumus 2.4	F1-Score	11



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	41
Lampiran 2	Formulir Bimbingan	42



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA