



**ANALISIS TOPIK DAN PERBANDINGAN KLASIFIKASI  
PADA KOLOM KOMENTAR VIDEO YOUTUBE EDUKASI  
INDONESIA MENGGUNAKAN PENDEKATAN LATENT  
DIRICHLET ALLOCATION**

**LAPORAN SKRIPSI**



**ALBERT  
41518120026**

**UNIVERSITAS  
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA JAKARTA  
2023**



**ANALISIS TOPIK DAN PERBANDINGAN KLASIFIKASI  
PADA KOLOM KOMENTAR VIDEO YOUTUBE EDUKASI  
INDONESIA MENGGUNAKAN PENDEKATAN LATENT  
DIRICHLET ALLOCATION**

**LAPORAN SKRIPSI**



**ALBERT  
41518120026**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana**

**UNIVERSITAS  
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA JAKARTA  
2023**

## LEMBAR PERNYATAAN KARYA SENDIRI

Yang bertanda tangan dibawah ini:

NIM : 41518120026

Nama : Albert

Judul Tugas Akhir : Analisis Topik dan Perbandingan Klasifikasi Pada Kolom  
Komentar Video YouTube Edukasi Indonesia  
Menggunakan Pendekatan Latent Dirichlet Allocation

Menyatakan bahwa, Laporan Tugas Akhir yang saya buat adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan unsur plagiat pada laporan Tugas Akhir saya, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 12 Januari 2023

UNIVERSITAS  
MERCU BUANA



Albert

## LEMBAR PENGESAHAN

Laporan Skripsi ini diajukan oleh:

NIM : 41518120026  
Nama Mahasiswa : Albert  
Judul Tugas Akhir : Analisis Topik Dan Perbandingan Klasifikasi Pada Kolom  
Komentar Video Youtube Edukasi Indonesia  
Menggunakan Pendekatan Latent Dirichlet Allocation

Tugas akhir ini telah disidangkan dihadapan Dewan Penguji dan diterima sebagai salah satu persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata-1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Disahkan oleh:

Pembimbing Tugas Akhir : Achmad Kodar, Drs. MT  
NIDN : 0323085801  
Ketua Penguji : Anis Cherid, SE., M.TI  
NIDN : 0328127203  
Penguji 1 : Harni Kusniyati, M.Kom  
NIDN : 0324068101  
Penguji 2 : Adi Hartanto, ST, M.Kom  
NIDN : 0717037202


Jakarta, 19 Januari 2023

Mengetahui,

Koordinator Tugas Akhir

Ketua Program Studi

  
(Wawan Gunawan, S.Kom., M.T.)

  
(Ir. Emil R. Kaburuan, Ph.D., IPM)

## KATA PENGANTAR

Puji dan syukur saya panjatkan atas kehadiran Tuhan Yang Maha Esa yang telah memberikan berkat yang berlimpah, sehingga saya dapat menyelesaikan laporan Tugas Akhir ini. Laporan ini disusun bertujuan untuk menjadi salah satu syarat kelulusan Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Universitas Mercu Buana.

Penulisan Laporan Tugas Akhir ini dihantarkan oleh dukungan dan bantuan secara moral dan serta pikiran dari berbagai pihak. Untuk itu, pada kesempatan ini, saya ingin memberi ucapan terima kasih kepada :

1. Bapak Achmad Kodar, Drs. MT, selaku Pembimbing Tugas Akhir yang telah membimbing dengan semua semangat, nasihat, dan ilmunya dalam penyusunan laporan tugas akhir ini.
2. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM. selaku Kepala Program Studi Teknik Informatika
3. Bapak Wawan Gunawan, S.Kom, MT. selaku Koordinator Tugas Akhir Teknik Informatika
4. Seluruh dosen dan staff akademi program Sarjana Strata (S1) Universitas Mercu Buana
5. Ibu dan teman-teman saya yang tidak ada hentinya dalam memberikan doa dan dukungan serta kasih sayang yang tiada hentinya.

Saya menyadari, bahwa laporan ini masih sangat jauh dari kata baik maupun sempurna. Oleh karena itu, sebelumnya saya ucapkan permintaan maaf saya yang setulus-tulusnya atas segala kekurangan dan keterbatasan dalam penulisan laporan kerja praktek ini. Semoga laporan tugas akhir ini dapat memberikan manfaat bagi pihak-pihak yang berkepentingan.

Jakarta, 13 Januari 2023

Albert

## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Albert  
NIM : 41518120026  
Judul Tugas Akhir : Analisis Topik dan Perbandingan Klasifikasi Pada Kolom  
Komentar Video YouTube Edukasi Indonesia  
Menggunakan Pendekatan Latent Dirichlet Allocation

Dengan ini, memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (None-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Ribet Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS  
MERCU BUANA

Jakarta, 12 Januari 2023



Albert

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>LEMBAR PERNYATAAN KARYA SENDIRI.....</b>	<b>ii</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR v</b>	
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>BAB I.....</b>	<b>1</b>
<b>PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
<b>BAB II .....</b>	<b>4</b>
<b>LANDASAN TEORI .....</b>	<b>4</b>
2.1 Penelitian Terdahulu .....	4
2.2 Dasar Teori .....	10
2.2.1 Topic Modelling .....	10
2.2.2 Latent Dirichlet Allocation .....	11
2.2.3 Topic Coherence .....	12

2.2.4 Analisa Sentimen .....	12
2.2.5. Support Vector Machine .....	13
2.2.6. Decision Tree .....	14
2.2.7 Preprocessing Text .....	15
2.2.8 TF-IDF Weighting (Term Frequency Inverse Document Frequency) .	15
2.2.9 Scraping .....	16
<b>BAB III.....</b>	<b>18</b>
<b>METODOLOGI PENELITIAN .....</b>	<b>18</b>
3.1 Karakteristik Penelitian .....	18
3.2 Sumber Pengambilan Data .....	18
3.3 Tahapan Penelitian .....	19
<b>BAB IV .....</b>	<b>21</b>
<b>HASIL DAN PEMBAHASAN .....</b>	<b>21</b>
4.1 Dataset .....	21
4.1.1 Proses Scraping Data .....	21
4.1.2 Hasil Scraping Data .....	22
4.1.3 Informasi Data .....	23
4.2 Pre-processing .....	25
4.2.1 Case Folding .....	25
4.2.2 Tokenizing .....	26
4.2.3 Filtering.....	26
4.2.4 Stemming.....	26
4.3 Pembuatan Model.....	27
4.3.1 Pemisahan Data Training dan Data Testing .....	27
4.3.2 Proses Vektorisasi.....	27
4.3.3 Model Support Vector Machine .....	28



4.3.4 Model Decision Tree .....	29
4.3.5 Pembuatan Model Latent Dirichlet Allocation.....	29
4.4 Visualisasi Data.....	31
4.4.1 Jumlah Sentimen.....	31
4.4.2 Permodelan Topik.....	33
4.5 Pengujian .....	33
4.5.1 Perhitungan Klasifikasi Teks Data .....	33
4.5.2 Pengujian Latent Dirichlet Allocation .....	34
4.6 Evaluasi Data.....	34
4.6.1 Analisa Sentimen dengan SVM dan Decision Tree .....	34
4.6.2 Hasil Topik dengan LDA .....	40
4.6.3 Rangkuman hasil penelitian.....	46
<b>BAB V .....</b>	<b>47</b>
<b>PENUTUP.....</b>	<b>47</b>
5.1 Kesimpulan.....	47
5.2 Saran.....	47
<b>DAFTAR PUSTAKA.....</b>	<b>48</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>50</b>
Lampiran Bimbingan.....	50
SURAT PENGALIHAN HAK CIPTA.....	51
HALAMAN PERNYATAAN LUARAN TUGAS AKHIR.....	52
SURAT PERNYATAAN.....	53
Lampiran Kartu Identitas Mahasiswa .....	54
Lampiran Bukti Submit.....	55
Curriculum Vitae.....	56
Lampiran Naskah Jurnal.....	57

## DAFTAR GAMBAR

[Gambar 1.1 Pengguna Internet Per. Feb 2022 .....	1]
Gambar 2.1 Penjelasan Mengenai Topic Modelling.....	11
Gambar 2.2 Visualisasi Mengenai Latent Dirichlet Allocation.....	12
Gambar 2.3 Visualisasi Mengenai Analisa Sentimen.....	13
Gambar 2.4 Visualisasi Mengenai Support Vector Machine.....	13
Gambar 2.5 Visualisasi mengenai Decision Tree .....	14
Gambar 2.6 Persamaan TF-IDF .....	15
Gambar 2.7 Visualisasi Mengenai Tahapan Web Scraping.....	16
Gambar 3.1 Flowchart Penelitian.....	19
Gambar 4.1 Halaman Google Developer - YouTube API.....	20
Gambar 4.2 Kode JSON yang dikeluarkan untuk ditransformasi ke Python.....	21
Gambar 4.3 Hasil Scraping Data yang berupa csv .....	22
Gambar 4.4 Hasil output DataFrame hasil scraping pada channel Alam Semenit	22
Gambar 4.5 Hasil output DataFrame hasil scraping pada channel Satu Persen ...	23
Gambar 4.6 Hasil output DataFrame hasil scraping pada channel Kok Bisa? .....	23
Gambar 4.7 Proses pemisahan data training dan testing.....	26
Gambar 4.8 Proses Vektorisasi data training dan testing .....	27
Gambar 4.9 Memunculkan output dari corpus yang menjadi token .....	27
Gambar 4.10 Data training yang sudah tervektorisasi .....	27
Gambar 4.11 Proses Pembuatan Model Support Vector Machine .....	28
Gambar 4.12 Proses pembuatan model decision tree .....	28
Gambar 4.13 Proses pembuatan corpus untuk topic modelling.....	29
Gambar 4.14 Hasil term frequency dari sebuah corpus berdasarkan dictionary ..	29
Gambar 4.15 term frequency dengan bentuk dictionary .....	29
Gambar 4.16 Pembuatan topic modelling dengan Latent Dirichlet Allocation....	30
Gambar 4.17 Jumlah Sentimen dari Dataset Komentar channel Satu Persen.....	30
Gambar 4.18 Jumlah Sentimen dari Dataset Komentar channel Alam Semenit ..	31
Gambar 4.19 Jumlah Sentimen dari Dataset Komentar channel Kok Bisa?.....	31
Gambar 4.20 Contoh Visualisasi dengan WordCloud.....	32
Gambar 4.21 Contoh topik yang dihasilkan oleh LDA .....	33

Gambar 4.22 Penjelasan mengenai confusion matrix .....	34
Gambar 4.23 Perhitungan nilai akurasi .....	34
Gambar 4.24 Perhitungan dataset Satu Persen menggunakan SVM dengan dua percobaan .....	35
Gambar 4.26 Perhitungan dataset Satu Persen menggunakan Decision Tree dengan dua percobaan .....	35
Gambar 4.28 Perhitungan dataset Kok Bisa? menggunakan Support Vector Machine dengan dua percobaan .....	36
Gambar 4.30 Perhitungan dataset Kok Bisa? menggunakan Decision Tree dengan dua percobaan .....	36
Gambar 4.32 Perhitungan dataset Alam Semenit menggunakan Support Vector Machine dengan dua percobaan .....	37
Gambar 4.34 Perhitungan dataset Alam Semenit menggunakan Decision Tree dengan dua percobaan .....	37
Gambar 4.36 Hasil Topik dari dataset Satu Persen .....	38
Gambar 4.37 Hasil Nilai Coherence dan Jumlah Topik dari dataset Satu Persen	39
Gambar 4.38 Hasil Topik dari Dataset Alam Semenit .....	39
Gambar 4.39 Hasil Nilai Coherence dan Jumlah Topik Dataset Alam Semenit ..	40
Gambar 4.40 Hasil Topik dari Dataset Kok Bisa? .....	40
Gambar 4.41 Hasil Nilai Coherence dan Jumlah Topik dari Dataset Kok Bisa? .	41

UNIVERSITAS  
MERCU BUANA

## DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya.....	23
Tabel 3.1 Rincian mengenai sumber data .....	31
Tabel 4.1 Jumlah label positif dan negatif dari setiap kanal.....	37
Tabel 4.2 Hasil Casefolding.....	37
Tabel 4.3 Hasil tokenizing .....	38
Tabel 4.4 Hasil filtering .....	38
Tabel 4.5 Hasil stemming .....	39
Tabel 4.6 Hasil dan parameter terbaik berdasarkan tingkat akurasi .....	58
Tabel 4.7 Hasil terbaik berdasarkan model topic coherence .....	58

