

**KLASTERISASI TOPIK KONTEN CHANNEL YOUTUBE
OTOMOTIF INDONESIA 2024 MENGGUNAKAN
*LATENT DIRICHLET ALLOCATION***



SKRIPSI

MIFTAH WISNU RAMDANI

1803040090

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
AGUSTUS 2024**

**KLASTERISASI TOPIK KONTEN CHANNEL YOUTUBE
OTOMOTIF INDONESIA 2024 MENGGUNAKAN
*LATENT DIRICHLET ALLOCATION***



SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar sarjana komputer

MIFTAH WISNU RAMDANI

1803040090

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
AGUSTUS 2024**

HALAMAN PERSETUJUAN

Skripsi yang diajukan oleh :

Nama : Miftah Wisnu Ramdani

NIM : 1803040090

Prodi : Teknik Informatika

Fakultas : Teknik dan Sains

Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

Judul : Klasterisasi Topik konten Channel YouTube Otomotif Indonesia 2024 Menggunakan Latent Dirichlet Allocation

telah disetujui untuk diajukan dalam ujian skripsi

Purwokerto, 14 Agustus 2024

Pembimbing,



Abid Yanuar Badharudin, S.Kom., M.Kom.

NIK 2160528



HALAMAN PENGESAHAN

Skripsi yang diajukan oleh:

Nama : Miftah Wisnu Ramdani

NIM : 1803040090

Program Studi : Teknik Informatika

Fakultas : Teknik dan Sains

Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

Judul : Klasterisasi Topik konten *Channel YouTube*
Otomotif Indonesia 2024 Menggunakan Latent
Dirichlet Allocation

DEWAN PENGUJI

Penguji 1 (Pembimbing1) : Abid Yanuar Badharudin, S.Kom., M.Kom.

Penguji 2 : Achmad Fauzan, S.Kom., M.Cs.

Penguji 3 : Maulida Ayu Fitriani, S.Kom., M.Cs.

Ditetapkan di : Purwokerto

Tanggal :

Mengetahui

Dekan Fakultas Teknik dan Sains



HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan dibawah ini:

Nama : Miftah Wisnu Ramdani
NIM : 1803040090
Program Studi : Teknik Informatika
Fakultas : Teknik dan Sains
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

menyatakan dengan sebenar-benarnya bahwa skripsi ini adalah hasil karya saya dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta bukan hasil penjiplakan dari karya orang lain.

Demikian pernyataan ini saya buat dan apabila kelak dikemudian hari terbukti ada unsur penjiplakan, saya bersedia mempertanggungjawabkan sesuai dengan ketentuan yang berlaku.

Purwokerto, 14 Agustus 2024
Yang membuat pernyataan



Miftah Wisnu Ramdani
1803040090

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Purwokerto dan demi pengembangan ilmu pengetahuan, saya yang bertanda tangan di bawah ini:

Nama : Miftah Wisnu Ramdani
NIM : 1803040090
Program Studi : Teknik Informatika
Fakultas : Teknik dan Sains
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto
Jenis Karya : Skripsi

Menyetujui untuk memberikan Hak Bebas Royalti Noneksklusif (*non-exclusive Royalty-Free Right*) kepada Universitas Muhammadiyah Purwokerto atas karya ilmiah saya yang berjudul:

KLASTERISASI TOPIK KONTEN *CHANNEL YOUTUBE* OTOMOTIF
INDONESIA 2024 MENGGUNAKAN *LATENT DIRICHLET ALLOCATION*

Beserta perangkat lunak yang ada (jika diperlukan), dengan Hak Bebas Royalti Noneksklusif ini Universitas Muhammadiyah Purwokerto berhak menyimpan, mengalihmedia/mengalih formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya dengan tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sebenarnya.

Purwokerto, 14 Agustus 2024
Yang membuat pernyataan



Miftah Wisnu Ramdani
1803040090

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati, serta rasa syukur terhadap Allah Taala yang telah memberikan rahmat dan nikmat-Nya, maka saya persembahkan Laporan Skripsi ini kepada:

1. Allah Taala yang senantiasa melimpahkan nikmat dan kasih sayang-Nya serta kelancaran dalam setiap proses pelaksanaan sehingga dapat menyelesaikan skripsi ini.
2. Kedua orang tua saya Bapak Nurisman dan Ibu Turkini yang senantiasa memberikan dukungan berupa moril, materil, bimbingan dan doanya untuk saya sampai saat ini.
3. Keluarga besar saya yang selalu mendoakan dan memberikan dukungan semangat.
4. Bapak Dr. Jebul Suroso, S.Kp., Ns., M.Kep. sebagai Rektor Universitas Muhammadiyah Purwokerto.
5. Bapak Dr. T. Ir. Iskahar, S.T., M.T., sebagai Dekan Fakultas Teknik dan Sains (FTS), Universitas Muhammadiyah Purwokerto (UMP).
6. Bapak Agung Purwo Wicaksonoo, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Universitas Muhammadiyah Purwokerto.
7. Bapak Abid Yanuar Badharudin, S.Kom., M.Kom. selaku Dosen Pembimbing1 Skripsi yang telah memberi arahan, bimbingan dan petunjuk dengan penuh kesabaran.
8. Bapak Achmad Fauzan, S.Kom., M.Cs. Selaku penguji 1 yang telah memberi arahan, bimbingan dan petunjuk dengan penuh kesabaran.
9. Ibu Maulida Ayu Fitriani, S.Kom., M.Cs. Selaku penguji 2 yang telah memberi arahan, bimbingan dan petunjuk dengan penuh kesabaran.
10. Terimakasih juga kepada teman Teknik Informatika dari tahun angkatan 2018 hingga saat ini yang telah memberi dukungan.

Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu menyelesaikan laporan tugas akhir ini

HALAMAN MOTTO

"Dan barang siapa yang bertakwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya." (Q.S At-Talaq: 4)



KATA PENGANTAR

Puji syukur atas kehadiran Allah Taala yang telah melimpahkan rahmat, hidayah dan anugerah-Nya sehingga laporan skripsi ini dapat diselesaikan sesuai waktu yang telah ditentukan. Adapun judul skripsi ini adalah **“KLASTERISASI TOPIK KONTEN *CHANNEL YOUTUBE OTOMOTIF INDONESIA 2024 MENGGUNAKAN LATENT DIRICHLET ALLOCATION*”**. Skripsi ini diajukan untuk memenuhi syarat kelengkapan kurikulum pada program studi Teknik Informatika Fakultas Teknik dan Sains Universitas Muhammadiyah Purwokerto. Pembuatan laporan ini mengacu dan berpedoman pada bahan kuliah, petunjuk dari dosen pembimbing skripsi, referensi dan literatur yang terkait dengan penulisan laporan.

Laporan ini jauh dari kata sempurna dan masih banyak kekurangan mengingat keterbatasan pengalaman dan kemampuan, oleh karena itu kritik dan saran yang membangun sangat diharapkan demi hasil yang lebih baik dimasa mendatang. Akhirnya, besar harapan agar kehadiran laporan ini dapat memberikan manfaat yang berarti untuk para pembaca, dan yang terpenting adalah semoga dapat turut serta memajukan ilmu pengetahuan.

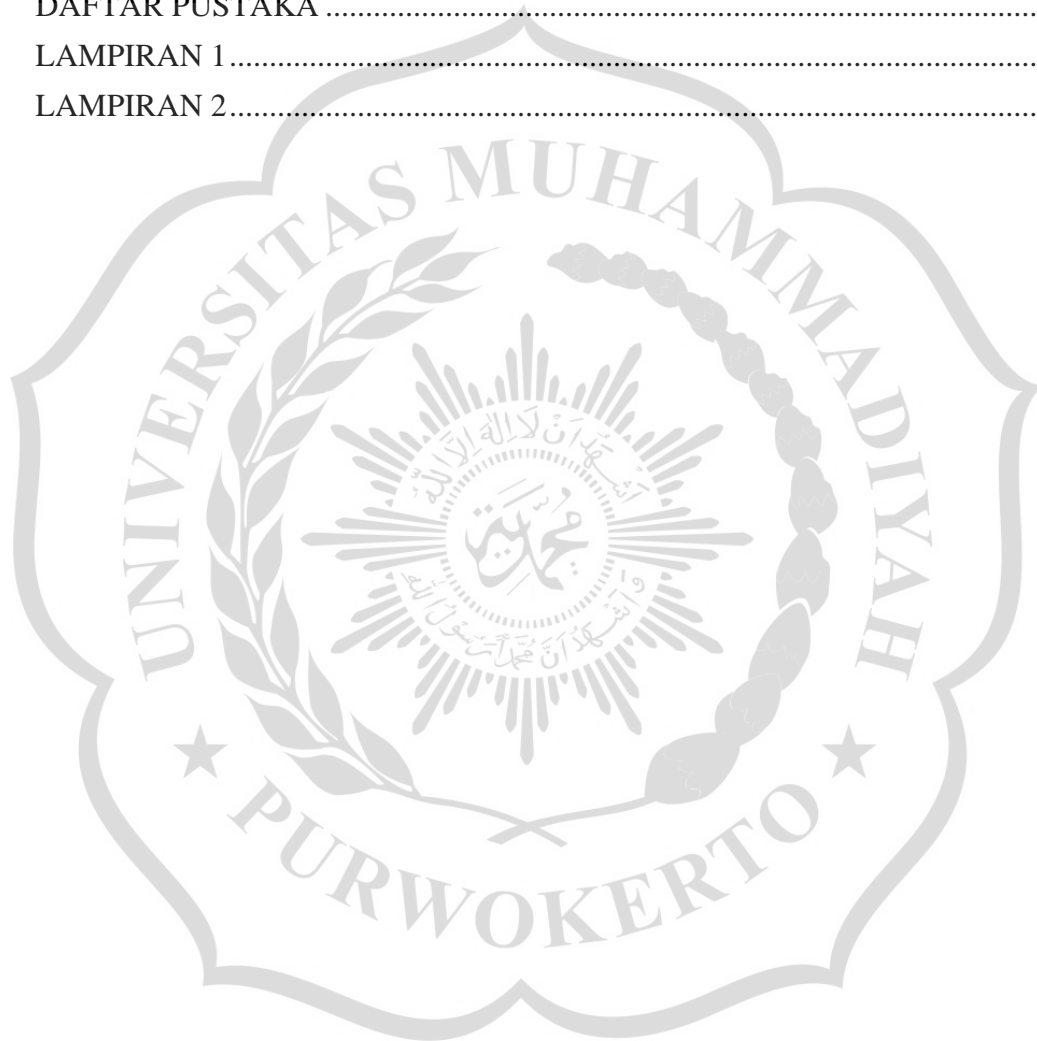
Purwokerto, 14 Agustus 2024
Yang membuat pernyataan

Miftah Wisnu Ramdani
1803040090

DAFTAR ISI

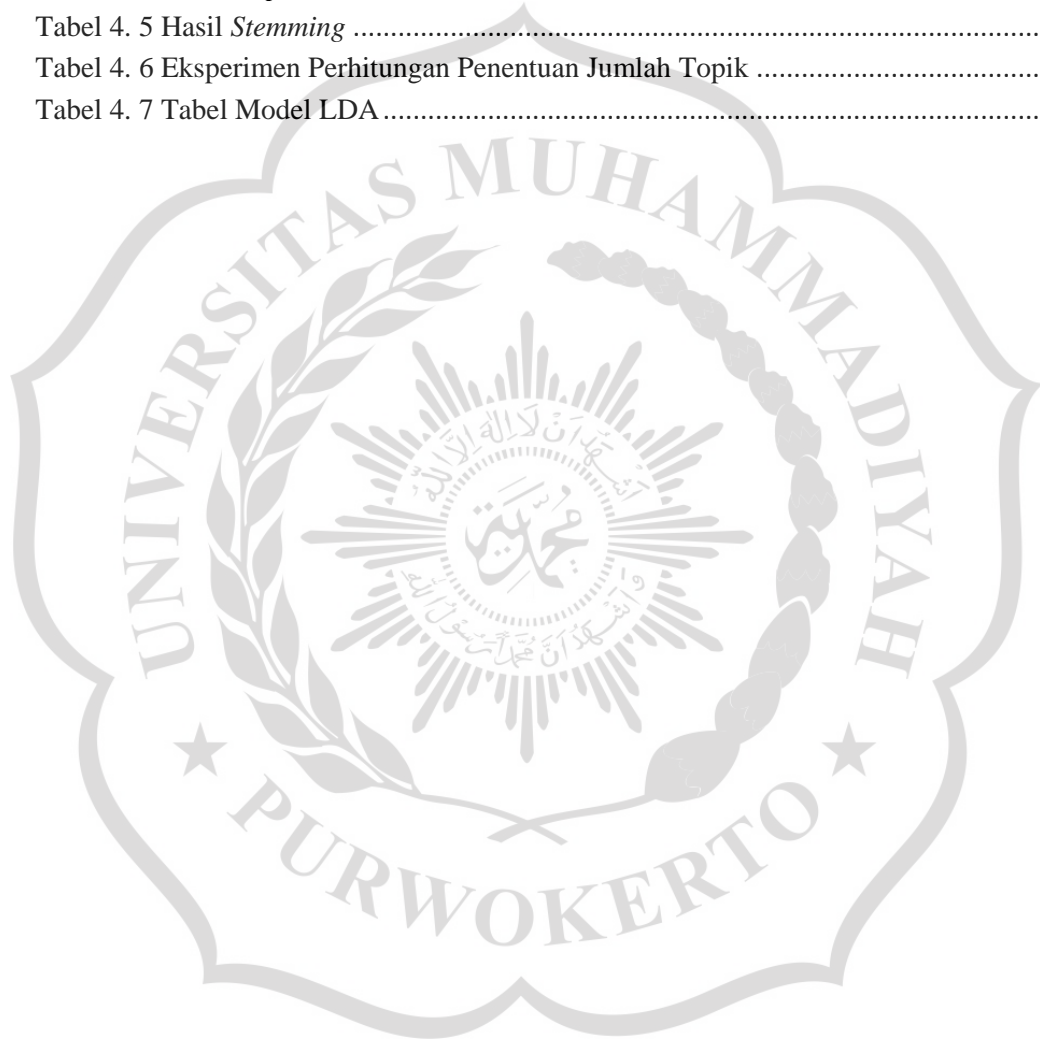
HALAMAN SAMPUL.....	i
HALAMAN JUDUL.....	i
HALAMAN PERSETUJIAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
ABSTRAK.....	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Perumusan Masalah	3
C. Batasan Masalah.....	3
D. Tujuan Penelitian	4
E. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
A. Peneliti Terdahulu	5
B. Landasan Teori.....	6
BAB III METODE PENELITIAN.....	13
A. Jenis Penelitian.....	13
B. Waktu dan Tempat	13
C. Metode Pengumpulan Data	13
D. Metode Analisis Data	14
BAB IV HASIL DAN PEMBAHASAN	19
A. Pengumpulan Data	19

B. <i>Preprocessing</i>	22
C. Proses Topik Modeling LDA	31
D. Visualisasi Data.....	38
BAB V KESIMPULAN DAN SARAN	48
A. Kesimpulan	48
B. Saran	49
DAFTAR PUSTAKA	50
LAMPIRAN 1	53
LAMPIRAN 2	59



DAFTAR TABEL

Tabel 2.1 Peneliti Terdahulu	5
Tabel 4. 1 Jumlah Nama Channel <i>YouTube</i> Berkategori Otomotif.....	19
Tabel 4. 2 Hasil <i>Cleaning</i>	24
Tabel 4. 3 Hasil <i>Tokenizing</i>	26
Tabel 4. 4 Hasil <i>Stopword</i>	28
Tabel 4. 5 Hasil <i>Stemming</i>	30
Tabel 4. 6 Eksperimen Perhitungan Penentuan Jumlah Topik	35
Tabel 4. 7 Tabel Model LDA	37

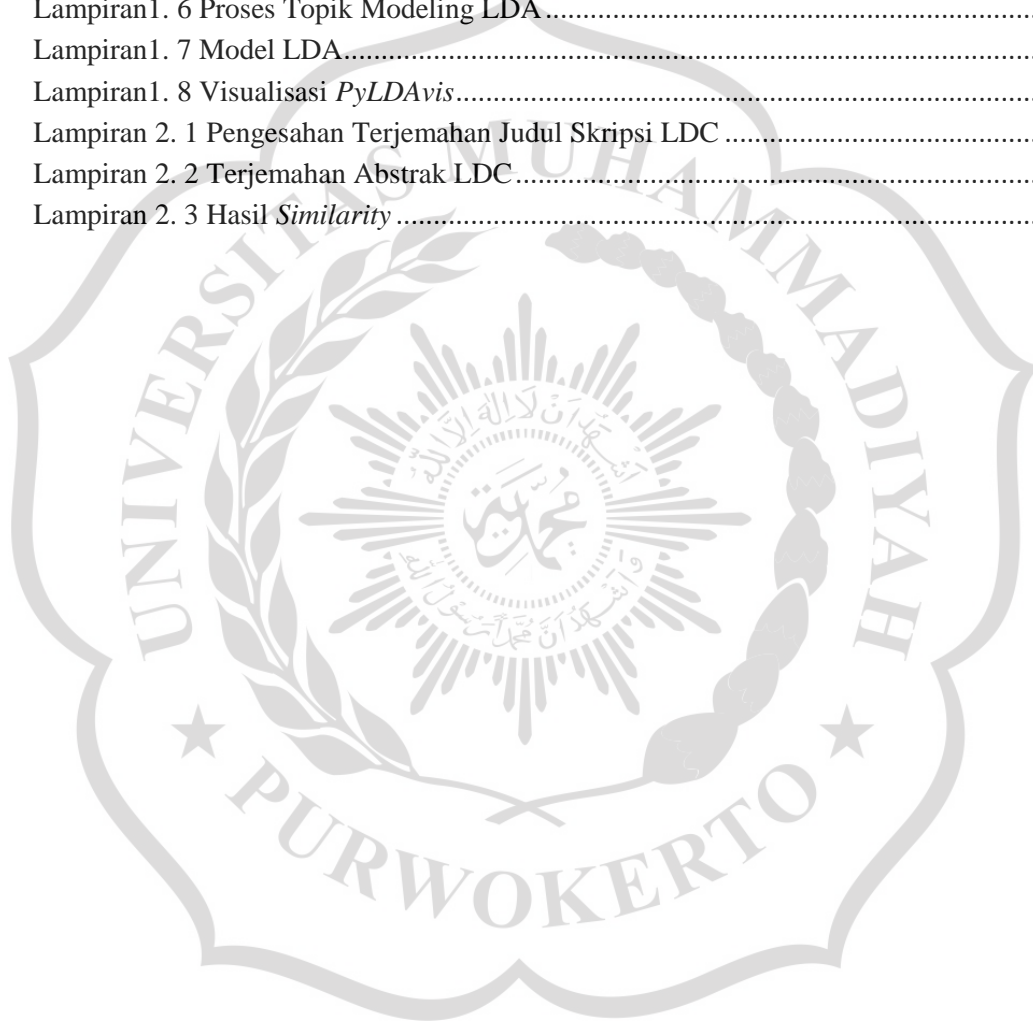


DAFTAR GAMBAR

Gambar 2.1 Diagram Model LDA (Zou et al., 2022)	11
Gambar 3. 1 <i>Flowchart</i> Alur Penelitian	14
Gambar 4. 1 Modul <i>googleapiclient</i>	20
Gambar 4. 2 <i>Source Code</i> Link <i>YouTube</i>	20
Gambar 4. 3 <i>Source Code</i> Menambahkan <i>List Channel</i> , Judul, dan <i>Subscriber</i>	21
Gambar 4. 4 <i>Source Code</i> Menyimpan Hasil <i>Web scraping</i>	21
Gambar 4. 5 <i>Source Code</i> menampilkan hasil <i>Web Scraping</i>	21
Gambar 4. 6 Hasil Proses <i>Scraping</i> data	22
Gambar 4. 7 <i>Source Code</i> Proses <i>Cleaning</i>	23
Gambar 4. 8 <i>Source Code</i> <i>Library NLTK</i>	25
Gambar 4. 9 <i>Source Code</i> Proses <i>Tokenizing</i>	25
Gambar 4. 10 <i>Source Code</i> Proses <i>Stopword</i>	27
Gambar 4. 11 <i>Source Code</i> Proses <i>Stemming</i>	29
Gambar 4. 12 <i>Source Code</i> Pemrosesan Kata <i>Stemming</i>	29
Gambar 4. 13 <i>Source Code</i> Modul Perhitungan LDA.....	32
Gambar 4. 14 <i>Source Code</i> input Hasil Data <i>Preprocessing</i>	32
Gambar 4. 15 <i>Source Code</i> Pembuatan <i>Corpus</i> Pada Dokumen	33
Gambar 4. 16 <i>Source Code</i> Percobaan Perhitungan 6 Topik.....	33
Gambar 4. 17 <i>Source Code</i> Perhitungan <i>Coherence Score</i>	33
Gambar 4. 18 <i>Source Code</i> Perhitungan <i>Perplexity Score</i>	34
Gambar 4. 19 <i>Source Code</i> Pengecekan Nilai Terbaik.....	34
Gambar 4. 20 <i>Source Code</i> Hasil <i>Perplexity Score</i> Dan <i>Coherence Score</i>	34
Gambar 4. 21 <i>Source Code</i> Grafik <i>Perplexity Score</i> dan <i>Coherence Score</i>	35
Gambar 4. 22 <i>Perplexity Score</i> dan <i>Coherence Score</i>	36
Gambar 4. 23 <i>Source Code</i> Model LDA	37
Gambar 4. 24 <i>Source Code</i> <i>PyLDAvis</i>	38
Gambar 4. 25 Visualisasi <i>PyLDAvis</i> Topik ke-1	39
Gambar 4. 26 Visualisasi <i>PyLDAvis</i> Topik Ke-2	41
Gambar 4. 27 Visualisasi <i>PyLDAvis</i> Topik Ke-3	42
Gambar 4. 28 Visualisasi <i>PyLDAvis</i> Topik ke-4	43
Gambar 4. 29 Visualisasi <i>PyLDAvis</i> Topik Ke-5	45
Gambar 4. 30 Visualisasi <i>PyLDAvis</i> Topik Ke-6	47

DAFTAR LAMPIRAN

Lampiran1. 1 <i>Scraping</i>	53
Lampiran1. 2 <i>Cleaning</i>	54
Lampiran1. 3 <i>Tokenizing</i>	55
Lampiran1. 4 <i>Stopword</i>	55
Lampiran1. 5 <i>Stemming</i>	56
Lampiran1. 6 Proses Topik Modeling LDA.....	56
Lampiran1. 7 Model LDA.....	58
Lampiran1. 8 Visualisasi <i>PyLDAvis</i>	58
Lampiran 2. 1 Pengesahan Terjemahan Judul Skripsi LDC	59
Lampiran 2. 2 Terjemahan Abstrak LDC.....	60
Lampiran 2. 3 Hasil <i>Similarity</i>	61



ABSTRAK

Perkembangan teknologi informasi saat ini menyebabkan perubahan dalam beberapa bidang kehidupan manusia, seperti dalam bidang otomotif, banyak orang menggunakan internet dan media sosial untuk memberikan pendapat dan memberikan informasi, *Youtube* merupakan media sosial yang digunakan untuk mengupload video, menonton berbagai video, Berdasarkan uraian di atas, penelitian ini menawarkan solusi untuk melakukan analisis topik modeling terhadap konten yang dibuat oleh para youtuber otomotif, Penelitian ini dilakukan dengan menggunakan salah satu metode topic modeling yaitu metode *Latent Dirichlet Allocation* (LDA) yang digunakan untuk membantu menentukan topik yang muncul pada akun *youtuber* otomotif. LDA ini dilakukan dengan meringkas, melakukan klasterisasi, menghubungkan hingga memproses data agar menghasilkan daftar topik yang diberi bobot untuk setiap dokumen, Proses pengumpulan data pada penelitian ini dilakukan dengan cara mengamati dan web scraping pada *YouTube* menggunakan *YouTube API V3*. Data yang di hasilkan pada tahap sebelumnya masih berupa data mentah, maka dari itu pada tahap selanjutnya akan dilakukan preprocessing sehingga nantinya akan di dapatkan data yang bersih agar memudahkan untuk proses selanjutnya. proses topik modeling menggunakan *Latent Dirichlet Allocation* bertujuan untuk membentuk topik model yang paling baik. Di hasilkan *perplexity score*: -8.1616 dan *coherence Score*: 0.4732 dari enam topik, dan topik paling dominan ada di topik ke satu dengan akurasi nilai 27,7 % dengan kata *honda*, *yamaha*, *review*, *kawasaki*, mesin, ninja, *nmax*, *vario*. Pada penelitian ini tentu tidak terlepas dari kekurangan dan kelemahan. Oleh karena itu peneliti menyarankan untuk penelitian selanjutnya adalah berikut 1. Menambahkan wordcloud untuk visualisasikan data. 2. Menambah beberapa algoritma *machine learning* yang lain agar dapat menjadi pembandingan algoritma mana yang lebih baik untuk penelitian selanjutnya.

Kata kunci: Klasterisasi, *YouTube*, Otomotif, Topik Modeling, *Latent Dirichlet Allocation*.

ABSTRACT

The current advancement of information technology has led to changes in various aspects of human life, including the automotive sector. Many people use the internet and social media to share opinions and information, with YouTube being a popular platform for uploading videos and watching a wide range of content. Based on this context, this study offers a solution to analyze the topic modeling of content created by automotive YouTubers. The research employs the Latent Dirichlet Allocation (LDA) method, one of the topic modeling techniques, to help identify the topics that emerge on automotive YouTuber accounts. The LDA method involves summarizing, clustering, linking, and processing data to generate a weighted list of topics for each document. The data collection in this study was conducted through observation and web scraping on YouTube using the YouTube API V3. The data obtained in the initial stage were raw, so the next step involved preprocessing to clean the data, making it easier for subsequent processing. The topic modeling process using Latent Dirichlet Allocation aimed to form the best possible topic model. The results showed a perplexity score of -8.1616 and a coherence score of 0.4732 from six topics, with the most dominant topic being the first topic, which had an accuracy rate of 27.7%, containing keywords such as Honda, Yamaha, review, Kawasaki, engine, Ninja, Nmax, and Vario. This study acknowledges certain limitations and weaknesses. Therefore, the researchers recommend the following for future studies: 1. Adding a word cloud to visualize the data, and 2. Incorporating additional machine learning algorithms to compare and determine which algorithm performs better in future research. Keywords: Clustering, YouTube, Automotive, Topic Modeling, Latent Dirichlet Allocation.

Keywords: Clustering, YouTube, Automotive, Topic Modeling, Latent Dirichlet Allocation.