



SKRIPSI

**PENGUNAAN TOPIC MODELING DENGAN
METODE *LATENT DIRICHLET ALLOCATION*
TERHADAP BERITA SAHAM INDEKS LQ45**

AKMAL ALIFFANDHI ANWAR

NPM 20081010148

DOSEN PEMBIMBING

Dr. RR. Ani Dijah Rahajoe, ST., M.Cs.

Dr. Ir. Kartini, S.Kom., MT

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

**PENGUNAAN TOPIC MODELING DENGAN
METODE *LATENT DIRICHLET ALLOCATION*
TERHADAP BERITA SAHAM INDEKS LQ45**

AKMAL ALIFFANDHI ANWAR

NPM 20081010148

DOSEN PEMBIMBING

Dr. RR. Ani Dijah Rahajoe, ST., M.Cs.

Dr. Ir. Kartini, S.Kom., MT

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongi

LEMBAR PENGESAHAN

PENGUNAAN TOPIC MODELING DENGAN METODE LATENT DIRICHLET ALLOCATION TERHADAP BERITA SAHAM INDEKS LQ45

Oleh :
AKMAL ALIFFANDHI ANWAR
NPM. 20081010148

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada tanggal 11 Maret 2025.

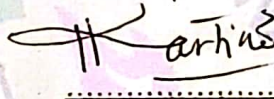
Menyetujui,

Dr. RR Ani Dijah Rahajoe, ST., M.Cs.
NIP. 19730512 200501 2003



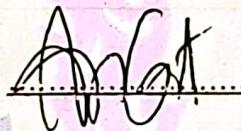
(Pembimbing I)

Dr. Ir. Kartini, S.Kom., MT.
NIP. 19611110 199103 2 001



(Pembimbing II)

Chrystia Aji Putra, S.Kom, M.T
NIP. 19861008 202121 1 001



(Ketua Penguji)

Eka Prakarsa Mandyartha, S.T., M.Kom
NIP. 19880525 201803 1 001



(Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongi

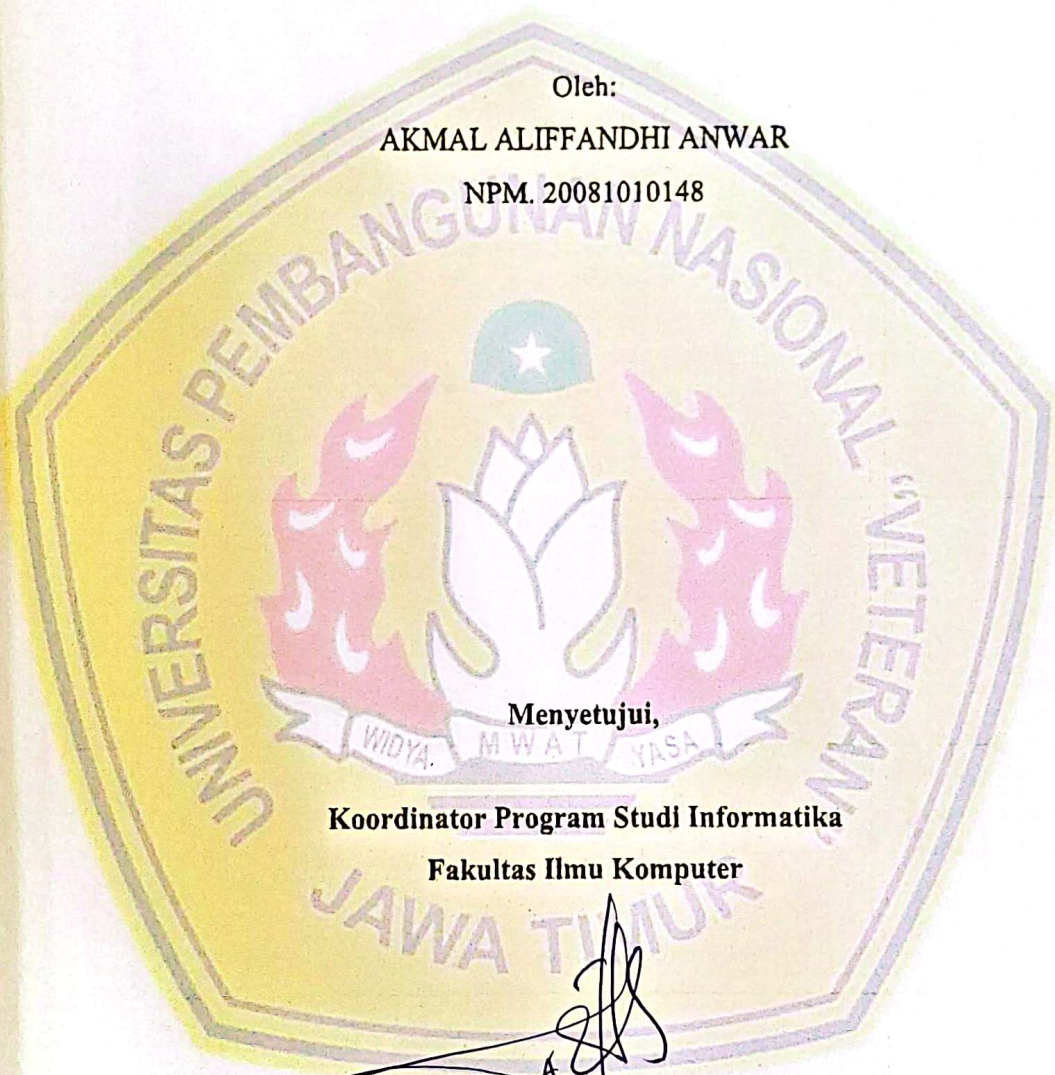
LEMBAR PERSETUJUAN

PENGUNAAN TOPIC MODELING DENGAN METODE *LATENT*
DIRICHLET ALLOCATION TERHADAP BERITA SAHAM INDEKS LQ45

Oleh:

AKMAL ALIFFANDHI ANWAR

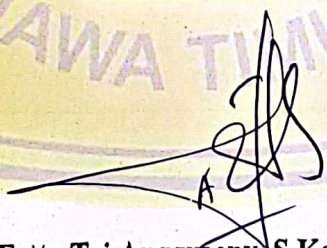
NPM. 20081010148



Menyetujui,

Koordinator Program Studi Informatika

Fakultas Ilmu Komputer


Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 2021212 005

Halaman ini sengaja dikosongi

SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa / NPM : Akmal Aliffandhi Anwar / 20081010148
Program Studi : Informatika
Dosen Pembimbing : 1. Dr. RR Ani Dijah Rahajoe, ST.,M.Cs.
2. Dr. Ir. Kartini, S.Kom., MT.

Dengan ini menyatakan bahwa Skripsi dengan judul "**PENGUNAAN TOPIC MODELING DENGAN METODE *LATENT DIRICHLET ALLOCATION* TERHADAP BERITA SAHAM INDEKS LQ45**" adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Universitas Pembangunan Nasional "Veteran" Jawa Timur.



Surabaya, 11 Maret 2025

Mahasiswa,

Akmal Aliffandhi Anwar

NPM. 20081010148

Halaman ini sengaja dikosongi

ABSTRAK

Nama Mahasiswa/NPM : Akmal Aliffandhi Anwar / 20081010148

Judul Skripsi : Penggunaan Topic Modeling dengan Metode *Latent Dirichlet Allocation* terhadap Berita Saham Indeks LQ45

Dosen Pembimbing : 1. Dr. RR. Ani Dijah Rahajoe, ST., M.Cs.
2. Dr. Ir. Kartini, S.Kom., MT

Investasi saham di Indonesia menjadi pilihan menarik karena potensi keuntungan tinggi meskipun memiliki risiko besar. Pandemi COVID-19 berdampak pada penurunan IHSG. Namun hal ini justru meningkatkan jumlah investor, terutama generasi milenial dan Gen Z, berkat kemudahan teknologi dalam transaksi saham. Analisis informasi menjadi kunci untuk pengambilan keputusan investasi, tetapi tantangan muncul dari volume dan kompleksitas berita yang terus berkembang. Teknik *topic modeling* seperti *Latent dirichlet allocation* (LDA) digunakan untuk mengatasi tantangan ini, memungkinkan identifikasi tema utama secara otomatis dan efisien. Penelitian ini menerapkan LDA pada berita saham LQ45 dengan tahapan pra-pemrosesan data, pembagian data latih dan uji, serta evaluasi coherence score. Hasilnya menunjukkan lima topik utama, termasuk "Analisis Teknikal Saham" dan "Kinerja Keuangan Bank," dengan coherence score tertinggi 0.66238 pada lima topik. Meski performa LDA lebih rendah dibandingkan HDP dan NMF, metode ini efektif mempercepat analisis dan memahami dinamika pasar saham secara terstruktur.

Kata Kunci: Topic Modeling, Berita Saham, Latent Dirichlet Allocation

Halaman ini sengaja dikosongi

ABSTRACT

Student Name/NPM : Akmal Aliffandhi Anwar / 20081010148
Thesis Title : Utilization of Topic Modeling with the *Latent dirichlet allocation* Method on LQ45 Index Stock News
Advisors : 1. Dr. RR. Ani Dijah Rahajoe, ST., M.Cs.
2. Dr. Ir. Kartini, S.Kom., MT

Stock investment in Indonesia has become an attractive choice due to its high potential returns despite significant risks. The COVID-19 pandemic impacted the decline of the Jakarta Composite Index (JCI) but increased the number of investors, especially millennials and Gen Z, driven by the ease of technology in stock transactions. Information analysis is crucial for investment decision-making, but challenges arise from the growing volume and complexity of news. Topic modeling techniques such as *Latent dirichlet allocation* (LDA) are used to address these challenges, enabling automatic and efficient identification of main themes. This study applies LDA to LQ45 stock news through data preprocessing, training and testing dataset splitting, and coherence score evaluation. The results reveal five main topics, including "Technical Stock Analysis" and "Bank Financial Performance," with the highest coherence score of 0.66238 on five topics. Although LDA performs lower than HDP and NMF, this method effectively accelerates analysis and provides a structured understanding of stock market dynamics.

Keywords: Topic Modeling, Stock News, Latent Dirichlet Allocation

Halaman ini sengaja dikosongi

KATA PENGANTAR

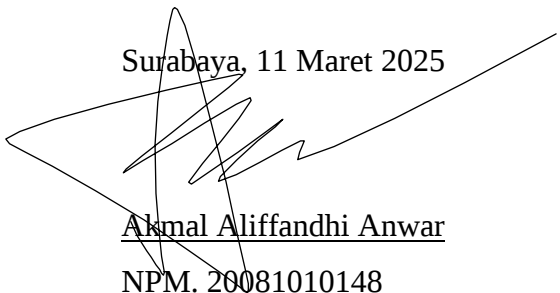
Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Penggunaan Topic Modeling dengan Metode *Latent Dirichlet Allocation* Terhadap Berita Saham Indeks LQ45”** dapat terselesaikan dengan baik. Penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU. selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Dr. RR Ani Dijah Rahajoe, ST.,M.Cs. selaku Dosen Pembimbing 1 dan Ibu Dr. Ir. Kartini, S.Kom., MT selaku dosen pembimbing 2, yang telah membantu dan memberikan arahan dalam menyelesaikan skripsi ini dengan baik.
5. Chrystia Aji Putra, S.Kom, M.T selaku Dosen Penguji pertama dan Eka Prakarsa Mandyartha, S.T., M.Kom. selaku Dosen Penguji kedua, yang telah memberi arahan serta saran kepada penulis sehingga penulis dapat menyelesaikan penyusunan skripsi ini.
6. Ketiga orangtua saya, yang telah maupun selalu memberikan motivasi, cinta, semangat, dan doa sehingga penulis dapat menyelesaikan pendidikan sampai masa perkuliahan selesai.
7. Om Yusron dan Tante Lis, yang telah menggantikan posisi sebagai orangtua selama di masa perantauan.
8. Keluarga besar Himatifa 2022-2023 yang telah memberikan pengalaman menyenangkan di tengah masa perkuliahan.

9. Keluarga besar Himatifa 2023-2024 yang telah menjadi sumber inspirasi berharga, memberikan pelajaran mendalam tentang tanggung jawab dan makna dari seorang pemimpin.
10. Tarisa Kamilia sebagai *support system* terbesar selama masa penyusunan dan penulisan skripsi.
11. Kevin, Hanan, dan Arya yang telah memberikan dukungan baik secara emosional maupun menggendong *rank* di *game mobile legends* selama masa penulisan skripsi.
12. Farhan, Ferdi, Rifki, Dani, dan Irsyad yang telah bersedia menjadi tempat berkeluh kesah, memberikan motivasi, dan menggendong di game *valorant* selama masa penulisan skripsi.
13. Serta yang terakhir kepada seluruh pihak yang penulis tidak dapat sebutkan satu persatu.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 11 Maret 2025



Akmal Aliffandhi Anwar

NPM. 20081010148

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN ORISINALITAS	v
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL	xix
DAFTAR KODE	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar belakang.....	1
1.2. Rumusan Masalah	4
1.3. Batasan masalah.....	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA.....	7
2.1. Penelitian Terdahulu	7
2.2. Dasar Teori.....	9
2.2.1. Berita Saham	9
2.2.2. Saham Indeks LQ45.....	10
2.2.3. Kontan.id.....	10
2.2.4. <i>Scraping</i>	11
2.2.5. <i>Machine Learning</i>	11
2.2.6. <i>Data Mining</i>	12
2.2.7. <i>Natural language Processing (NLP)</i>	13

2.2.8. Pemodelan Topik (<i>Topic Modeling</i>).....	13
2.2.9. <i>Bag of Words</i> (BoW)	14
2.2.10. <i>Gibbs Sampling</i>	15
2.2.11. <i>Latent Dirichlet Allocation</i> (LDA)	16
2.2.12. <i>Hiearchial Dirichlet Process</i> (HDP)	18
2.2.13. <i>Non-Negative Matrix Factorization</i> (NMF)	20
2.2.14. <i>Coherence Score</i>	22
2.2.15. <i>Word Cloud</i>	23
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....	25
3.1. Metode Penelitian	25
3.1.1. Pengumpulan Data.....	25
3.1.2. Pra-pemrosesan Data	27
3.1.3. Penerapan Topik Modeling dengan LDA.....	35
3.1.4. Evaluasi	36
3.2. Desain Sistem	37
3.2.1. Pengumpulan Data.....	37
3.2.2. Pra-pemrosesan Data	39
3.2.3. Penerapan Topic Modeling dengan LDA.....	41
3.2.4. Evaluasi	48
BAB IV PENGUJIAN DAN ANALISA	51
4.1. Pengumpulan Data.....	51
4.2. Persiapan.....	53
4.2.1. Pembersihan Data	53
4.2.2. <i>Case Folding</i>	54
4.2.3. <i>Cleansing Data</i>	55
4.2.4. <i>Tokenizing</i>	57

4.2.5. <i>Stopwords Removal</i>	58
4.2.6. <i>Stemming</i>	59
4.2.7. <i>Pembagian Data</i>	61
4.3. <i>Implementasi dan Pengujian Model</i>	62
4.3.1 <i>Pembuatan Dictionary dan Bag of Words</i>	62
4.3.2 <i>Implementasi Latent Dirichlet Allocation (LDA)</i>	64
4.3.3 <i>Implementasi Hierarchial Dirichlet Process (HDP)</i>	67
4.3.4 <i>Implementasi Non-Negative Matrix Factorization (NMF)</i>	70
4.4. <i>Perbandingan dan Analisis Model</i>	72
4.5. <i>Output dan Visualisasi</i>	76
4.6. <i>Implementasi graphical user interface GUI</i>	81
BAB V KESIMPULAN DAN SARAN	85
5.1 <i>Kesimpulan</i>	85
5.2 <i>Saran</i>	86
DAFTAR PUSTAKA	89

Halaman ini sengaja dikosongi

DAFTAR GAMBAR

Gambar 2.1 <i>Plate Notation Metode LDA</i>	17
Gambar 2.2 <i>Plate Notation HDP</i>	19
Gambar 2.3 <i>Skema Chinese Restaurant Franchise</i>	20
Gambar 2.4 <i>Proses Dekomposisi Matriks pada NMF</i>	22
Gambar 3.1 <i>Alur Penelitian</i>	25
Gambar 3.2 <i>Flowchart Pengumpulan Data</i>	26
Gambar 3.3 <i>Flowchart Pre-processing Data</i>	28
Gambar 3.4 <i>Flowchart Case Folding Dataset</i>	29
Gambar 3.5 <i>Flowchart Cleansing Dataset</i>	30
Gambar 3.6 <i>Flowchart Tokenizing Dataset</i>	31
Gambar 3.7 <i>Flowchart Stopword Removal pada Dataset</i>	32
Gambar 3.8 <i>Flowchart Stemming pada Dataset</i>	34
Gambar 3.9 <i>Flowchart Penerapan Model LDA untuk Topic Modeling</i>	35
Gambar 3.10 <i>Flowchart Gibbs Sampling</i>	35
Gambar 4.1 <i>Dataframe Hasil Scraping</i>	53
Gambar 4.2 <i>Rekap Data sebelum dan sesudah Dilakukan Pembersihan</i>	54
Gambar 4.3 <i>Data setelah Proses Case Folding</i>	55
Gambar 4.4 <i>Data setelah Proses Cleansing Data</i>	56
Gambar 4.5 <i>Data setelah Proses Tokenizing</i>	57
Gambar 4.6 <i>Word cloud teks sebelum dan sesudah stopwords removal</i>	59
Gambar 4.7 <i>Word cloud Teks Sebelum dan Sesudah Stemming</i>	61
Gambar 4.8 <i>Distribusi Kata dalam Data Latih dan Data Uji</i>	62
Gambar 4.9 <i>Hasil Bag of Words (BoW)</i>	64
Gambar 4.10 <i>Grafik Perbandingan Coherence Score Metode</i>	73

Gambar 4.11 Visualisasi Topik 1.....	77
Gambar 4.12 Visualisasi Topik 2.....	78
Gambar 4.13 Visualisasi Topik 3.....	79
Gambar 4.14 Visualisasi Topik 4.....	80
Gambar 4.15 Tampilan Awal.....	81
Gambar 4.16 Upload File.....	82
Gambar 4.17 Input Jumlah Topik	82
Gambar 4.18 Output GUI Model	84

DAFTAR TABEL

Tabel 3.1 Contoh Data Hasil <i>Scrapping</i> Portal Berita	38
Tabel 3.2 Contoh Data Hasil <i>Case Folding</i>	39
Tabel 3.3 Contoh Data Hasil <i>Cleansing</i>	40
Tabel 3.4 Contoh Data Hasil <i>Tokenizing</i>	40
Tabel 3.5 Contoh Data Hasil <i>Stopwords Removal</i>	40
Tabel 3.6 Contoh Data Hasil <i>Stemming</i>	41
Tabel 3.7 Contoh Data Hasil <i>Pre-processing</i>	42
Tabel 3.8 Contoh Inisialisasi Topik per Kata.....	42
Tabel 3.9 Jumlah Kata pada Setiap Topik (BoW).....	43
Tabel 3.10 Kata Yang Muncul di Setiap Topik dalam Setiap Dokumen	44
Tabel 3.11 Distribusi Kata dalam Topik	44
Tabel 3.12 Distribusi Topik dalam Dokumen	45
Tabel 3.13 Probabilitas Kata	46
Tabel 3.14 Normalisasi Distribusi Kata dalam Topik.....	47
Tabel 4.1 Data Hasil <i>Scrapping</i> Portal Berita	51
Tabel 4.2 Tabel Hasil Evaluasi Model	73
Tabel 4.3 Tabel Statistik Deskriptif Evaluasi Model	74
Tabel 4.4 <i>Topic-Word Distribution</i> Model LDA	76
Tabel 4.5 Label Topik	80

Halaman ini sengaja dikosongi

DAFTAR KODE

Kode 4.1 Impor <i>Library</i> yang Diperlukan	52
Kode 4.2 Upload File Data Hasil <i>Scraping</i> ke <i>Notebook</i>	52
Kode 4.3 Menghapus Data Kosong dan Pengambilan Kolom ' <i>content</i> '	53
Kode 4.4 Proses <i>Case Folding</i> pada Dataset	54
Kode 4.5 Proses <i>Cleansing</i> pada Dataset	56
Kode 4.6 Proses <i>Tokenizing</i> pada Dataset	57
Kode 4.7 Proses <i>Stopwords Removal</i> pada Dataset.....	58
Kode 4.8 Proses <i>Stemming</i> pada Dataset	60
Kode 4.9 Pembagian Data Uji dan Data Latih	62
Kode 4.10 Pembuatan <i>Bag of Words</i> (BoW).....	63
Kode 4.11 Evaluasi <i>Coherence Score</i> Model LDA.....	65
Kode 4.12 Pembangunan dan Mencetak Hasil Model LDA	66
Kode 4.13 Evaluasi <i>Coherence Score</i> Model HDP	68
Kode 4.14 Pembangunan dan Mencetak hasil HDP	69
Kode 4.15 Evaluasi <i>Coherence Score</i> Model NMF	71
Kode 4.16 Pembangunan dan Mencetak Hasil NMF	72

Halaman ini sengaja dikosongi