



Penggunaan *Latent Dirichlet Allocation* untuk *Topic Modeling* Berita Saham Indeks LQ45

Akmal Aliffandhi Anwar¹, Ani Dijah Rahajoe^{*2}, Kartini³

^{1,2,3} Informatika, Fakultas Ilmu Komputer, UPN Veteran Jawa Timur, Surabaya, Indonesia
Email: ¹akmalaanwar@gmail.com; ^{*2}anidijah.if@upnjatim.ac.id; ³kartini.if@upnjatim.ac.id

Anwar. A. A., Rahajoe. A. D., & Kartika. K. (2026). Penggunaan Latent Dirichlet Allocation untuk Topic Modeling Berita Saham Indeks LQ45. *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 12(2), 199-206

DOI: <https://doi.org/10.33050/cerita.v12i2.3805>

ABSTRAK

Investasi saham di Indonesia semakin diminati karena potensi keuntungan yang tinggi, meskipun memiliki risiko signifikan. Pandemi COVID-19 menyebabkan penurunan IHSG, tetapi jumlah investor, terutama milenial dan Gen Z, meningkat berkat kemudahan teknologi dalam transaksi saham. Analisis informasi menjadi krusial dalam pengambilan keputusan investasi, namun meningkatnya volume dan kompleksitas berita menimbulkan tantangan. Untuk mengatasi hal ini, metode Latent Dirichlet Allocation (LDA) digunakan untuk mengidentifikasi tema utama secara otomatis dan efisien. Evaluasi coherence score pada rentang dua hingga sepuluh topik menunjukkan bahwa jumlah topik optimal adalah empat, dengan nilai tertinggi sebesar 0,60495. Selain itu, coherence score menunjukkan LDA memiliki performa lebih tinggi dibandingkan HDP dan NMF. Dengan rata-rata tertinggi (0,559308) dan median terbesar (0,581115), LDA mampu menghasilkan topik yang lebih berkualitas dan konsisten. Meskipun standar deviasinya lebih besar dibanding HDP, keterkaitan kata yang lebih koheren menjadikannya metode paling efektif dalam memahami pola utama dalam berita saham.

Kata kunci: Topic Modeling, Berita Saham, Latent Dirichlet Allocation (LDA)

ABSTRACT

Stock investment in Indonesia has become increasingly attractive due to its high potential returns, despite significant risks. The COVID-19 pandemic led to a decline in the Jakarta Composite Index (JCI), but the number of investors, particularly millennials and Gen Z, increased due to the ease of technology in stock transactions. Information analysis is crucial in investment decision-making; however, the growing volume and complexity of news present challenges. To address this, the Latent Dirichlet Allocation (LDA) method is used to automatically and efficiently identify key themes. Coherence score evaluation within a range of two to ten topics indicates that the optimal number is five, with the highest score of 0.604946 and the lowest of 0.390450 at two topics. Furthermore, the coherence score shows

that LDA outperforms HDP and NMF. With the highest mean (0.559308) and the largest median (0.581115), LDA generates higher-quality and more consistent topics. Although its standard deviation is larger than HDP, its more coherent word associations make it the most effective method for understanding key patterns in stock market news.

Keywords: *Topic Modeling, Stock News, Latent Dirichlet Allocation (LDA)*

I. PENDAHULUAN

Investasi saham merupakan salah satu instrumen investasi yang memiliki karakteristik high risk–high return, yaitu potensi keuntungan yang tinggi diikuti dengan tingkat risiko yang tinggi akibat fluktuasi harga saham. Oleh karena itu, investasi saham menjadi salah satu pilihan yang semakin diminati masyarakat Indonesia karena berpotensi memberikan keuntungan dalam waktu yang relatif singkat (Farild et al., 2023).

Pandemi COVID-19 berdampak buruk pada Indonesia, termasuk pasar saham, yang ditandai dengan penurunan IHSG. Laporan BPS (2020) pada tahun 2020 menunjukkan IHSG turun 5,1% dibanding tahun sebelumnya (BPS Provinsi DKI Jakarta, 2021). Namun, kondisi ini justru dimanfaatkan investor untuk berinvestasi di pasar saham. Data KSEI mencatat peningkatan SID hingga 7,5 juta, naik 92,99% sejak 2019 (KSEI, 2022). Hal ini dipengaruhi oleh kemajuan teknologi yang berperan dalam mempermudah transaksi saham (Toto, 2022).

Investor membutuhkan informasi yang akurat untuk membuat keputusan yang tepat. Semakin banyak informasi yang tersedia, semakin kuat keyakinan investor dalam menanamkan modal pada suatu perusahaan (Sari et al., 2019). Dalam praktiknya, berita finansial merupakan sumber informasi tidak terstruktur yang memuat berbagai peristiwa ekonomi, politik, dan bisnis yang dapat memengaruhi persepsi investor maupun pergerakan harga saham. Oleh karena itu, analisis berita secara manual menjadi kurang efisien ketika jumlah dokumen semakin besar sehingga diperlukan pendekatan otomatis berbasis Natural Language Processing (NLP) untuk mengekstraksi informasi yang relevan dari kumpulan berita finansial. Salah satu teknik yang banyak digunakan adalah topic modeling, yaitu pendekatan dalam NLP yang bertujuan mengidentifikasi tema utama dalam sekumpulan dokumen (García-Méndez et al. 2023; Chen et al. 2023).

Topic modeling merupakan teknik unsupervised learning dalam bidang Natural Language Processing (NLP) yang bertujuan mengidentifikasi struktur topik laten dari sekumpulan dokumen berdasarkan pola kemunculan kata. Setiap dokumen direpresentasikan sebagai kombinasi beberapa topik, sedangkan setiap topik direpresentasikan sebagai distribusi probabilitas terhadap kata-kata penyusunnya (Blei et al., 2003). Salah satu metode yang paling banyak digunakan untuk topic modeling adalah Latent Dirichlet Allocation (LDA), yaitu model probabilistik yang mampu mengekstraksi struktur tematik dari korpus teks berskala besar secara otomatis. Metode ini efektif diterapkan pada data teks dalam jumlah besar karena mampu menghasilkan topik yang mudah diinterpretasikan tanpa memerlukan data berlabel (Wiranto & Mila Rosyida Uswatunnisa, 2022; Wang, 2026; García-Méndez et al. 2023).

Penelitian Pavithra dan Savitha (2024) menunjukkan bahwa model LDA menghasilkan kualitas topik yang lebih baik dibandingkan HDP dan NMF berdasarkan nilai coherence score pada konfigurasi tertentu. Namun demikian, performa model tetap dipengaruhi oleh pemilihan jumlah topik sehingga optimasi parameter menjadi tahapan yang penting dalam proses topic modeling. (Pavithra & Savitha, 2024).

Dalam penelitian yang berjudul “Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali,” ditemukan bahwa LDA dapat digunakan untuk mengelompokkan dokumen berbahasa Bali dengan tingkat akurasi mencapai 62% ketika setiap dokumen direpresentasikan oleh 3.000 kata. Selain itu, kesamaan antar kata serta jumlah kata yang dipilih dalam dokumen sangat memengaruhi akurasi proses klasterisasi (Sanjaya, 2021). Meskipun penelitian tersebut menunjukkan kemampuan LDA dalam mengelompokkan dokumen, objek penelitian yang digunakan masih berupa teks umum sehingga karakteristik berita finansial yang memiliki istilah teknis dan dinamika pasar belum banyak dikaji.

Meskipun berbagai penelitian telah menunjukkan bahwa metode Latent Dirichlet Allocation (LDA) mampu mengidentifikasi struktur topik laten secara efektif pada berbagai jenis korpus teks, penerapannya pada berita pasar saham Indonesia, khususnya yang berkaitan dengan indeks LQ45, masih relatif terbatas. Karakteristik berita finansial berbeda dengan dokumen umum karena mengandung informasi ekonomi, sentimen pasar, serta istilah teknis yang dapat memengaruhi perilaku investor dan dinamika harga saham. Penelitian terkini juga menunjukkan bahwa analisis berita finansial berbasis Natural Language Processing (NLP) dan topic modeling mampu menghasilkan informasi tematik yang bermanfaat untuk memahami kondisi pasar serta mendukung proses pengambilan keputusan investasi. Oleh karena itu, diperlukan penelitian yang tidak hanya mengevaluasi performa model topic modeling, tetapi juga menginterpretasikan topik yang dihasilkan sesuai dengan konteks pasar saham Indonesia (García-Méndez et al. 2023; Groß-Klußmann, 2024; Bybee et al. 2024).

Penelitian ini bertujuan untuk menerapkan metode Latent Dirichlet Allocation (LDA) dalam analisis topic modeling pada berita saham yang termasuk dalam indeks LQ45. Dengan penelitian ini, diharapkan dapat mempermudah analisis saham yang terdaftar dalam indeks tersebut melalui identifikasi topik utama dari berita terkait saham LQ45.

Berdasarkan kesenjangan penelitian tersebut, penelitian ini memberikan tiga kontribusi utama dalam kajian text mining finansial. Pertama, penelitian ini memetakan struktur topik laten pada berita saham indeks LQ45 menggunakan pendekatan probabilistik berbasis LDA. Kedua, penelitian ini membandingkan performa LDA, Hierarchical Dirichlet Process (HDP), dan Non-Negative Matrix Factorization (NMF) menggunakan coherence score sebagai metrik evaluasi intrinsik untuk menentukan model terbaik. Ketiga, penelitian ini tidak hanya mengevaluasi performa model, tetapi juga menginterpretasikan setiap topik berdasarkan konteks pasar saham Indonesia sehingga menghasilkan pemahaman yang lebih komprehensif mengenai pola informasi yang membentuk dinamika pasar modal domestik. Dengan demikian, penelitian ini tidak hanya bersifat implementatif, tetapi juga memberikan kontribusi konseptual terhadap

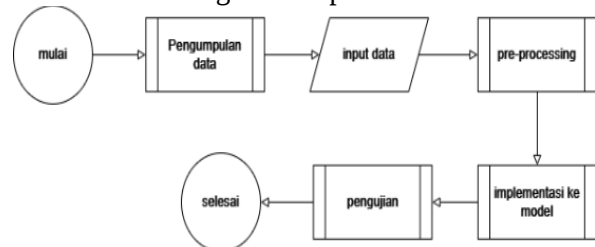
penerapan topic modeling pada analisis berita finansial.

Karena LDA merupakan metode unsupervised learning, penelitian ini berfokus pada proses eksplorasi struktur topik laten tanpa menggunakan label kelas sehingga hasil yang diperoleh sepenuhnya ditentukan oleh pola distribusi kata dalam korpus berita.

II. METODE PENELITIAN

Penelitian ini menggunakan metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik utama dalam berita saham yang berkaitan dengan indeks LQ45. LDA dipilih karena kemampuannya dalam memodelkan distribusi topik secara probabilistik pada korpus teks besar tanpa memerlukan label kategorisasi awal. Dibandingkan pendekatan berbasis clustering tradisional, LDA mampu menangkap struktur laten dokumen secara lebih fleksibel melalui distribusi Dirichlet.

Dalam konteks berita saham yang heterogen dan dinamis, pendekatan generatif ini lebih sesuai dibandingkan metode supervised yang memerlukan anotasi sentimen atau kategori topik secara manual. Alur penelitian ini akan disajikan dalam bentuk diagram alir pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Tabel 1. Contoh Dataset

Search	Judul	Link	Date	Content
lq45	Simak Top Gainers LQ45 di Tengah Pelemahan IHSG pada Kamis (15/8), Cek 3 Emiten Ini	https://investasi.kontan.co.id/news/simak-top-gainers-lq45-di-tengah-pelemahan-ihsg-pada-kamis-15-8-2024-1718-wib	Kamis , 15 Agust us 2024 / 17:18 WIB	Simak Top Gainers LQ45 di tengah...

Kata kunci yang digunakan dalam proses ini meliputi 'lq45', 'bbca', 'bbri', 'bbni', 'bmri', 'brisi', 'arto', dan 'bbtn', yang merupakan saham-saham utama di Indonesia, khususnya yang tergabung dalam indeks LQ45. Sebanyak 2.530 baris data berhasil dikumpulkan dan disimpan dalam format file CSV. Setelah seluruh data terkumpul, kolom 'content' diekstraksi untuk tahap pemrosesan selanjutnya.

B. Praproses Data (Text Preprocessing)

Proses ini dilakukan untuk membersihkan, menyederhanakan, dan menstrukturkan teks sehingga analisis topik menjadi lebih akurat dan efisien.

Tabel 2. Contoh Hasil Pre-processing Data

Sebelum		Sesudah	
Sepuluh	indeks	['sepuluh',	'indeks',
sektoral	menguat	'sektoral',	'kuat',
bersama	dengan	'sama', 'ihsg.',	'sektor',
IHSG.	Sektor	'kesehatan',	
kesehatan	melonjak	'melonjak.',	'sektor',
2,79%.	Sektor barang	'barang',	'konsumsi',
konsumsi	nonprimer	'nonprimer',	'melesat.']
melesat	2,58%.		

Tabel 2. Menampilkan hasil dari data yang telah melewati hasil pre-processing. Proses prapemrosesan mencakup case folding untuk standarisasi huruf, cleansing untuk menghapus karakter tidak relevan, tokenizing untuk memecah teks, stopword removal untuk menghilangkan kata umum, dan stemming untuk menyederhanakan kata, memastikan data lebih terstruktur dan siap untuk analisis topic modeling.

C. Pembagian Data

Karena topic modeling merupakan pendekatan unsupervised learning, seluruh korpus digunakan untuk membangun model tanpa pemisahan data latih dan data uji. Evaluasi kualitas model dilakukan menggunakan metrik coherence score yang mengukur konsistensi semantik antar kata dalam satu topik. Coherence score digunakan sebagai metrik intrinsik untuk mengevaluasi kualitas topik berdasarkan tingkat keterkaitan semantik antar kata dalam topik, bukan untuk mengukur akurasi klasifikasi.

D. Pembuatan Bag-of-Words

Setelah pembagian data, teks dikonversi ke bentuk numerik menggunakan Bag-of-Words (BoW), yang merepresentasikan kumpulan kata unik berdasarkan frekuensi tanpa

memperhatikan urutannya. Metode ini sederhana, efisien, dan sering digunakan dalam prapemrosesan teks karena kemudahan interpretasi serta perhitungannya.

Tabel 3. Contoh Hasil Representasi BoW

Kata	Dokumen 1	Dokumen 2
alam	0	1
asia	0	1
...		
tahun	1	0
tbk	1	0

Metode representasi ini mendukung algoritma seperti Latent Dirichlet Allocation (LDA) dalam mengidentifikasi topik melalui distribusi probabilitas kata dalam dokumen.

E. Pemodelan Topic Modeling dengan LDA

Pemodelan topik dilakukan menggunakan metode Latent Dirichlet Allocation (LDA) sebagaimana diperkenalkan oleh David M. Blei et al. (2003). LDA merupakan model generatif probabilitas yang mengasumsikan bahwa setiap dokumen terdiri atas distribusi beberapa topik laten, dan setiap topik merupakan distribusi probabilitas atas kata-kata dalam korpus.

Parameter yang digunakan dalam penelitian ini meliputi: Jumlah topik (k): diuji dalam rentang 2–10 topik; Alpha (α): parameter distribusi topik dalam dokumen; Beta/Eta (η): parameter distribusi kata dalam Topic Modeling; Jumlah iterasi: xxx iterasi; Random state 42 untuk menjaga replikasi eksperimen Minimum frequency threshold 5 untuk mengurangi noise vocabulary. Seluruh korpus digunakan dalam proses pemodelan karena topic modeling merupakan pendekatan unsupervised learning yang tidak memerlukan pembagian data latih dan data uji.

F. Evaluasi Model dan Pencarian Jumlah Topik Optimal

Proses *topic modeling* dengan *Latent Dirichlet Allocation* (LDA) diawali dengan pelatihan model menggunakan data yang telah diproses sebelumnya. Parameter utama yang digunakan merupakan jumlah topik, yang berkisar antara 2 hingga 10 topik. Hal ini dilakukan untuk menemukan jumlah topik optimal yang diukur dengan metrik *coherence score* yang nantinya akan dicetak. Untuk

memberikan analisis yang lebih komprehensif, LDA dibandingkan dengan dua metode pemodelan topik lainnya, yaitu *Hierarchical Dirichlet Process* (HDP) dan *Non-Negative Matrix Factorization* (NMF).

G. Pencetakan Topik oleh Model

Setelah evaluasi *coherence score* pada model dilakukan, tahap selanjutnya adalah menampilkan topik-topik yang dihasilkan oleh model LDA. Topik yang dicetak didasarkan pada jumlah topik dengan nilai *coherence score* tertinggi, guna memastikan relevansi dan kualitas tema yang diidentifikasi.

H. Karakteristik Model

Latent Dirichlet Allocation (LDA) merupakan model generatif probabilistik yang mengasumsikan bahwa setiap dokumen merupakan distribusi campuran dari beberapa topik laten, dan setiap topik merupakan distribusi probabilistik atas kata. Model ini tidak memerlukan label atau kategori awal, sehingga termasuk dalam pendekatan *unsupervised learning*.

III. HASIL DAN PEMBAHASAN

A. Evaluasi Model

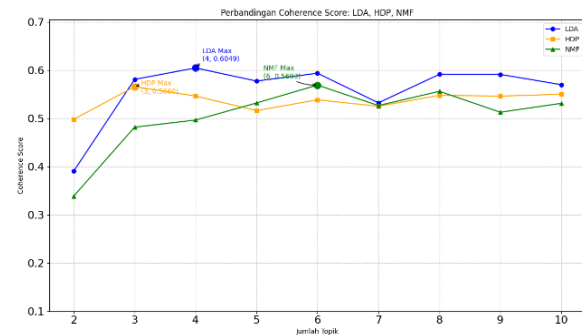
Berdasarkan perbandingan *coherence score*, model LDA menunjukkan kinerja yang paling tinggi dibandingkan model HDP dan NMF. Hasil evaluasi terhadap ketiga metode terdapat di tabel 4.

Tabel 4. Hasil *coherence score* ketiga model

Topics	LDA	HDP	NMF
2	0.39045	0.49816	0.33875
3	0.58112	0.56605	0.48194
4	0.60495	0.54652	0.4967
5	0.57764	0.51628	0.53217
6	0.59392	0.53868	0.56929
7	0.53248	0.52532	0.52675
8	0.5916	0.54806	0.55643
9	0.59149	0.54638	0.51292
10	0.57012	0.55031	0.5314

Kemudian data dalam tabel telah dirangkum dalam bentuk diagram garis untuk mempermudah analisis, sebagaimana

ditampilkan pada Gambar 2. Visualisasi ini membantu dalam mengidentifikasi tren serta membandingkan kinerja setiap metode pada berbagai jumlah topik. Diagram garis ini secara efektif menggambarkan fluktuasi *coherence score*, sehingga memudahkan observasi terhadap jumlah topik optimal untuk masing-masing model.



Gambar 2. Grafik Coherence Score

Berdasarkan hasil evaluasi, LDA menunjukkan performa terbaik pada empat topik dengan *coherence score* 0.60495, lebih tinggi dibandingkan HDP dan NMF pada jumlah topik tersebut. HDP memiliki *coherence score* yang lebih stabil, dengan nilai tertinggi 0.56605 pada tiga topik dan fluktuasi minimal dibanding metode lainnya. Sebaliknya, NMF memiliki nilai lebih rendah pada jumlah topik kecil.

LDA mulai mengalami penurunan setelah lebih dari empat topik, namun tetap lebih unggul dibandingkan HDP dan NMF. HDP mempertahankan kestabilannya dengan *coherence score* antara 0.49816 hingga 0.55031. Sementara itu, NMF menunjukkan peningkatan dengan nilai 0.56929 pada enam topik dan 0.55643 pada delapan topik. Secara keseluruhan, LDA unggul dalam koherensi topik, HDP lebih stabil, dan NMF meningkat pada topik lebih besar.

Tabel 5. Statistika Deskriptif Coherence Score

	LDA	HDP	NMF
Mean	0.559308	0.537307	0.50515
Standard Deviation	0.062823	0.01943	0.064108
Max Coherence	0.604946	0.566049	0.569295
Min Coherence	0.39045	0.498164	0.338753
Range	0.214496	0.067885	0.230542

LDA memiliki *coherence score* rata-rata tertinggi sebesar 0.559308, menunjukkan kualitas topik yang lebih baik dibandingkan HDP dan NMF. Meskipun demikian, standar deviasi LDA lebih besar (0.062823) dibandingkan HDP (0.019430), yang menunjukkan bahwa HDP memiliki stabilitas lebih tinggi.

NMF memiliki fluktuasi terbesar dengan *range* tertinggi 0.230542, sedangkan HDP paling stabil dengan *range* 0.067885. Secara keseluruhan, LDA unggul dalam koherensi topik, HDP lebih stabil, dan NMF memiliki variasi tinggi serta performa lebih rendah dibandingkan metode lainnya.

B. Hasil Topic Modeling

Setelah mengevaluasi *coherence score* model, langkah berikutnya adalah menghasilkan topik yang diperoleh dari model LDA. Proses ini dilakukan berdasarkan jumlah topik dengan *coherence score* paling tinggi. Berdasarkan Tabel 4, hasil terbaik diperoleh pada konfigurasi empat topik dengan nilai *coherence score* sebesar 0.60495. Oleh karena itu, model LDA dengan empat topik dipilih sebagai model terbaik untuk mengidentifikasi tema utama pada berita saham indeks LQ45.

Tabel 6. Hasil Model LDA

Topik	Hasil
T1	0.032*"bank" + 0.025*"tumbuh" + 0.018*"kredit" + 0.016*"tingkat" + 0.015*"yoy" + 0.010*"syariah" + 0.010*"btn" + 0.010*"biaya" + 0.009*"milik" + 0.007*"bris"
T2	0.027*"support" + 0.026*"kuat" + 0.024*"resistance" + 0.021*"potensi" + 0.017*"koreksi" + 0.017*"target" + 0.015*"area" + 0.015*"wave" + 0.013*"gerak" + 0.012*"gera"
T3	0.095*"kuat" + 0.054*"top" + 0.047*"sektoral" + 0.043*"lemah" + 0.026*"konsumsi" + 0.025*"gainers" + 0.024*"energi" + 0.024*"teknologi" + 0.024*"losers" + 0.024*"primer"
T4	0.036*"bank" + 0.035*"bbri" + 0.025*"bbca" + 0.025*"bmri" + 0.024*"goto" + 0.024*"arto" + 0.023*"kali" + 0.018*"buka" + 0.017*"turun" + 0.014*"hari"

Topik 1 membahas sektor perbankan dan pertumbuhan kredit sebagai indikator ekonomi utama. Kata dengan bobot tertinggi, seperti bank (0.032) dan tumbuh (0.025), menekankan pentingnya ekspansi kredit. Istilah tingkat

(0.016) dan yoy (0.015) menunjukkan analisis pertumbuhan tahunan. Selain itu, terdapat bahasan mengenai bank syariah dan peran Bank BTN dalam pembiayaan perumahan. Kata biaya (0.010) dan milik (0.009) mengacu pada efisiensi keuangan serta kepemilikan saham. Keseluruhan topik ini menyoroti perkembangan industri perbankan dan pengaruhnya terhadap ekonomi. Topik ini mencerminkan dominasi sektor perbankan dalam indeks LQ45, yang secara historis memiliki kapitalisasi pasar besar dan sensitivitas tinggi terhadap kebijakan moneter. Kemunculan istilah yoy dan tumbuh menunjukkan bahwa berita saham banyak dipengaruhi oleh publikasi kinerja keuangan kuartalan dan pertumbuhan kredit tahunan. Hal ini menegaskan bahwa sentimen pasar terhadap saham perbankan sangat terkait dengan indikator fundamental makroekonomi.

Dominasi istilah yang berkaitan dengan sektor perbankan menunjukkan bahwa pemberitaan mengenai saham LQ45 masih dipengaruhi oleh kondisi fundamental industri perbankan sebagai kontributor utama kapitalisasi pasar di Bursa Efek Indonesia. Hal ini menunjukkan bahwa informasi mengenai pertumbuhan kredit, laba, dan kinerja bank menjadi perhatian utama investor dalam mengambil keputusan investasi.

Topik 2 berfokus pada analisis teknikal saham, dengan istilah seperti support (0.027) dan resistance (0.024) yang menggambarkan batas harga dalam perdagangan saham. Kata kuat (0.026) dan potensi (0.021) menunjukkan evaluasi pergerakan saham, sementara koreksi (0.017) dan target (0.017) mengacu pada prediksi harga saham. Selain itu, konsep wave (0.015) menandakan penggunaan teori gelombang dalam analisis pasar. Topik ini mencerminkan strategi teknikal yang digunakan investor untuk memproyeksikan tren harga saham. Dominasi istilah support, resistance, dan wave menunjukkan bahwa pemberitaan saham LQ45 tidak hanya berorientasi pada fundamental perusahaan, tetapi juga kuat dipengaruhi pendekatan teknikal. Hal ini mencerminkan perilaku investor ritel yang cenderung menggunakan analisis grafik dalam pengambilan keputusan jangka pendek.

Keberadaan istilah support, resistance, target, dan wave menunjukkan bahwa media finansial tidak hanya menyampaikan informasi fundamental, tetapi juga banyak memuat analisis teknikal yang digunakan investor jangka pendek

dalam menentukan waktu beli maupun jual saham.

Topik 3 membahas performa pasar saham dan sektor unggulan. Kata kuat (0.095) menandakan analisis saham yang menunjukkan tren positif, sedangkan top (0.054) dan sektoral (0.047) mengacu pada sektor dominan. Sebaliknya, kata lemah (0.043) mencerminkan saham yang mengalami penurunan. Sektor seperti konsumsi (0.026), energi (0.024), dan teknologi (0.024) menjadi perhatian utama. Istilah gainers (0.025) dan losers (0.024) digunakan untuk menggambarkan saham yang mengalami kenaikan atau penurunan signifikan. Topik ini mengindikasikan adanya rotasi sektor dalam dinamika pasar. Kemunculan istilah gainers dan losers menunjukkan fokus media terhadap pergerakan harian indeks sektoral. Fenomena ini konsisten dengan karakter pasar saham Indonesia yang volatil dan responsif terhadap sentimen global maupun domestik.

Topik ini menggambarkan bahwa berita pasar modal juga dipengaruhi oleh rotasi sektor industri. Kemunculan sektor energi, konsumsi, dan teknologi menunjukkan adanya perubahan fokus investor terhadap sektor tertentu sesuai kondisi ekonomi maupun sentimen pasar.

Topik 4 membahas pergerakan saham perbankan besar serta emiten teknologi. Kata bank (0.036) menegaskan fokus utama pada sektor keuangan, sementara bbri (0.035), bbca (0.025), dan bmri (0.025) menunjukkan perhatian terhadap saham BRI, BCA, dan Mandiri. Kata goto (0.024) dan arto (0.024) menandakan keterlibatan saham teknologi seperti GoTo dan Bank Jago. Istilah kali (0.023) serta buka (0.018) mengacu pada momentum perdagangan, sementara turun (0.017) dan hari (0.014) mencerminkan analisis pergerakan harga saham. Munculnya saham seperti BCA, BRI, Mandiri serta emiten teknologi seperti GOTO menunjukkan bahwa perhatian pasar terkonsentrasi pada saham berkapitalisasi besar (big caps) dan saham berbasis ekonomi digital. Hal ini memperlihatkan struktur pasar LQ45 yang masih didominasi sektor keuangan, namun mulai terdampak transformasi digital.

Dominasi saham BBRI, BBKA, BMRI, serta GOTO menunjukkan bahwa pemberitaan pasar modal lebih banyak berfokus pada emiten dengan kapitalisasi pasar besar karena saham-saham tersebut memiliki likuiditas tinggi dan menjadi penggerak utama indeks LQ45.

Dari hasil *topic modeling* yang telah dianalisis, setiap topik memiliki karakteristik unik berdasarkan kata-kata dominannya. Untuk memahami lebih lanjut, seperti yang tersaji di tabel 7.

Tabel 7. Hasil Model LDA

Topik	Label
Topik 1	Perbankan dan Pertumbuhan Kredit
Topik 2	Analisis Teknikal dan Pergerakan Harga Saham
Topik 3	Kinerja Sektor Saham dan Pergerakan Pasar
Topik 4	Saham Perbankan dan Pergerakan Emiten Besar

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa metode Latent Dirichlet Allocation (LDA) efektif dalam mengidentifikasi struktur tematik pada berita saham indeks LQ45. Evaluasi coherence score pada rentang dua hingga sepuluh topik menunjukkan bahwa empat topik merupakan konfigurasi optimal dengan nilai tertinggi 0.60495.

Dibandingkan dengan HDP dan NMF, LDA menghasilkan rata-rata coherence tertinggi, meskipun dengan variasi yang lebih besar. Secara substantif, topik yang terbentuk merefleksikan karakteristik pasar saham Indonesia yang didominasi sektor perbankan, analisis teknikal, dinamika sektoral, dan konsentrasi pada emiten berkapitalisasi besar.

Temuan ini menunjukkan bahwa topic modeling dapat menjadi pendekatan eksploratif yang relevan dalam analisis text mining finansial untuk memahami pola informasi yang beredar di media pasar modal.

DAFTAR PUSTAKA

- BPS Provinsi DKI Jakarta. (2021, July 1). Indeks Harga Saham Gabungan Menurut Sektor dan Bulan, 2020. <https://Jakarta.Bps.Go.Id/Id/Statistics-Table/1/MTI3IzE=/Indeks-Harga-Saham-Gabungan-Menurut-Sektor-Dan-Bulan--2020>.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of*

- Machine Learning Research 3, 3, 993–1022. <https://doi.org/10.1016/B978-0-12-411519-4.00006-9>
- BYBEE, L., KELLY, B., MANELA, A. and XIU, D. (2024), Business News and Business Cycles. *J Finance*, 79: 3105-3147. <https://doi.org/10.1111/jofi.13377>
- Chen, W., Rabhi, F., Liao, W., & Al-Qudah, I. (2023). Leveraging State-of-the-Art Topic Modeling for News Impact Analysis on Financial Markets: A Comparative Study. *Electronics*, 12(12), 2605. <https://doi.org/10.3390/electronics12122605>
- Farild, M., Sawaji, M. I., & Poddala, P. (2023). Analisis teknikal sebagai dasar pengambilan keputusan dalam transaksi saham. *FORUM EKONOMI: Jurnal Ekonomi, Manajemen, Dan Akuntansi*, 25(4), 734–739.
- García-Méndez, S., de Arriba-Pérez, F., Barros-Vila, A. et al. Automatic detection of relevant information, predictions and forecasts in financial news through topic modelling with Latent Dirichlet Allocation. *Appl Intell* 53, 19610–19628 (2023). <https://doi.org/10.1007/s10489-023-04452-4>
- Groß-Klußmann, A. Learning deep news sentiment representations for macro-finance. *Digit Finance* 6, 341–377 (2024). <https://doi.org/10.1007/s42521-024-00107-2>
- KSEI. (2022). Berita Pers: Investor Pasar Modal Tembus 10 Juta.
- Pardede, D. L. C., & Waskita, M. A. I. (2023). Analisis Pemodelan Topik Untuk Ulasan Tentang Peduli Lindungi. *Jurnal Ilmiah Informatika Komputer*, 28(1), 17–26. <https://doi.org/10.35760/ik.2023.v28i1.7925>
- Pavithra, & Savitha. (2024). Topic Modeling for Evolving Textual Data Using LDA, HDP, NMF, BERTOPIC, and DTM With a Focus on Research Papers. *Journal of Technology and Informatics (JoTI)*, 5(2), 53–63. <https://doi.org/10.37802/joti.v5i2.618>
- Sanjaya, N. A. (2021). Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 127–134. <https://doi.org/10.25126/jtiik.202183556>
- Sari, W. W., Fathoni, A. F., & Haryetti. (2019). Pengaruh Informasi Kinerja Saham terhadap Pengambilan Keputusan Investasi dengan Herding Behavior sebagai Variabel Mediasi. *Jurnal Ekonomi Fakultas Ekonomi Dan Bisnis Universitas Riau*, 27(3), 282–292. <http://je.ejournal.unri.ac.id/>
- Toto, R. K. (2022). FAKTOR YANG MEMPENGARUHI MINAT GEN Z DALAM BERINVESTASI DI PASAR MODAL. *Science*, 7(1), 1–8. <http://link.springer.com/10.1007/s00232-014-9701-9>
<http://link.springer.com/10.1007/s00232-014-9700-x>
<http://dx.doi.org/10.1016/j.jmr.2008.11.017>
<http://linkinghub.elsevier.com/retrieve/pii/S1090780708003674>
<http://www.ncbi.nlm.nih.gov/pubmed/1191>
- Wiranto, W., & Mila Rosyida Uswatunnisa. (2022). Topic Modeling for Support Ticket using Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(6), 998–1005. <https://doi.org/10.29207/resti.v6i6.4542>
- Wang, X. (2026). Topic Modeling in Finance: A Review of Methods, Applications, and Challenges. *Journal of Risk and Financial Management*, 19(6), 399. <https://doi.org/10.3390/jrfm19060399>