

## TOPIK KONTRAS BERBASIS LIFT POSITIF-NEGATIF: STRATEGI BARU REKOMENDASI KONTEN

Tubagus M. Akhriza\*, Sigit Setyowibowo, Tubagus Arief Armanda dan  
Tubagus M Adrie Admira

STMIK PPKIA Pradnya Paramita (STIMATA)  
Jl. Laksda Adi Sucipto No.249a, Pandanwangi, Kec. Blimbing, Kota Malang  
Sekolah Tinggi Manajemen Informatika dan Komputer Jakarta STI&K  
Jalan BRI No. 17, Radio Dalam, Kebayoran B  
\*Corresponding Author

### ABSTRAK

Pemetaan topik tradisional seperti *Latent Dirichlet Allocation (LDA)* sering kali berfokus pada distribusi global kata tanpa memperhatikan hubungan spesifik antar topik. Penelitian ini memperkenalkan strategi baru dalam rekomendasi konten berbasis *Association Rule Learning (AR)* dengan *Lift* positif dan negatif. Metode ini tidak hanya mampu menemukan asosiasi yang sering muncul bersama (*Lift* positif), tetapi juga mengungkap kontras antara topik-topik yang jarang muncul bersama (*Lift* negatif), memberikan wawasan lebih dalam yang tidak dapat diidentifikasi oleh *LDA*. Pendekatan ini diterapkan pada dataset berita politik Indonesia untuk membangun sistem rekomendasi konten yang lebih cerdas dan dinamis, yang mampu menawarkan topik-topik kontras yang menarik minat pengguna serta memperkaya pengalaman mereka. Hasil eksperimen menunjukkan bahwa metode ini lebih fleksibel dalam mengidentifikasi keterkaitan dan anomali antar topik, sehingga memberikan rekomendasi yang lebih relevan dan terarah.

**Kata Kunci:** *Topik kontras, pemetaan topik, rekomendasi konten, Latent dirichlet allocation, Association rule.*

### PENDAHULUAN

Dalam era informasi yang serba cepat, pemetaan topik telah menjadi alat penting untuk mengelola dan memahami volume data teks yang terus meningkat, khususnya dalam konteks konten digital. Pemetaan topik memungkinkan identifikasi tema-tema utama yang terkandung dalam kumpulan teks, memfasilitasi berbagai aplikasi seperti rekomendasi konten, analisis tren, dan sistem pencarian informasi. Salah satu metode yang telah populer dalam pemetaan topik adalah *Latent Dirichlet Allocation (LDA)*, yang mengelompokkan kata-kata dalam dokumen menjadi topik-topik laten berdasarkan probabilitas kemunculan kata [1].

Keutamaan dari *LDA* adalah kemampuannya untuk secara efektif menemukan topik-topik tersembunyi di dalam kumpulan teks besar, yang menjadikannya metode pilihan untuk berbagai aplikasi mulai dari analisis sentimen hingga pemetaan topik politik; sedangkan keterbatasan metode ini adalah fokusnya pada distribusi probabilistik

global, yang sering kali mengabaikan hubungan spesifik antar topik, terutama dalam hal hubungan kontras atau negatif antar topik. *LDA* mengasumsikan bahwa topik-topik dalam dokumen memiliki distribusi campuran, tanpa menekankan korelasi negatif atau kejadian topik yang jarang muncul Bersama. Selain itu, evaluasi model ini cenderung mengandalkan metrik seperti perplexity, yang seringkali tidak sesuai dengan penilaian manusia terkait kualitas topik yang ditemukan [1]–[3].

*Neural Topic Models (NTM)* berbasis pembelajaran mesin telah mulai diperkenalkan dalam pemetaan topik, termasuk penggunaan *Graph Contrastive Topic Models (GCTM)*. *GCTM* memperkenalkan konsep pembelajaran kontras, di mana model dilatih untuk membedakan antara pasangan sampel positif dan negatif. Meskipun pendekatan ini lebih efektif dalam mengidentifikasi struktur topik laten, ia memiliki kelemahan signifikan: membutuhkan data latih yang cukup besar dan komputasi intensif untuk mengoptimalkan model [4], [5].

Dalam pendekatan pemetaan topik berbasis AI, algoritma LDA sering dikombinasikan dengan teknik semantik berbasis *knowledge base* untuk meningkatkan interpretasi topik yang dihasilkan. Salah satu metode yang digunakan adalah pengayaan topik dengan DBpedia Spotlight, di mana entitas diidentifikasi dalam teks dan dikaitkan dengan sumber data semantik dari Dbpedia. Sebuah publikasi menjelaskan penerapan LDA dan Dbpedia Spotline, menggunakan KNIME analytics platform[6].

Sebuah publikasi menggunakan *association rule (AR) learning* untuk topic modeling, yang fokus pada asosiasi positif untuk memperkuat koherensi topik [7]. Namun, potensi penerapan lift positif-negatif belum dikembangkan, di mana lift negatif bisa digunakan untuk memetakan topik kontras yang jarang muncul bersama, memberikan perspektif yang lebih dalam pada hubungan antar topik.

Sebagai alternatif, penggunaan AR *learning* dengan pengukuran lift menawarkan pendekatan berbeda dalam pemetaan topik. Tidak hanya menyoroti hubungan positif antar topik, tetapi juga mengidentifikasi topik-topik yang memiliki hubungan negatif—yaitu topik-topik yang jarang muncul bersamaan atau bahkan saling bertolak belakang. Dengan demikian, pendekatan ini memungkinkan penemuan kontras antar topik yang tidak dapat diungkapkan oleh metode berbasis distribusi probabilistik seperti LDA.

AR learning sudah digunakan secara luas dalam sistem rekomendasi (SR); seperti SR perpustakaan [8], [9], SR produk e-commerce dalam *data stream* transaksi[10], SR berbasis generative-AR untuk memprediksi produk berikutnya yang mungkin dibeli oleh pengguna e-commerce[11]. Metode-metode ini memanfaatkan nilai *lift* positif dalam merekomendasikan produk untuk dibeli bersama dengan produk utama yang pernah dilihat atau dibeli pengguna *e-commerce*.

Penelitian ini mengusulkan penggunaan AR *learning* dengan lift positif dan negatif untuk mengidentifikasi hubungan dan kontras antar topik dalam pemetaan topik. Aturan asosiasi yang

dihasilkan divisualisasikan menggunakan plot koordinat paralel, yang mampu menonjolkan perbedaan antar topik secara lebih jelas. Pendekatan ini diharapkan menawarkan perspektif baru, terutama dalam sistem rekomendasi konten, di mana topik yang bertolak belakang dengan arus utama dapat menarik perhatian pengguna.

Tujuan utama penelitian adalah mengeksplorasi potensi lift negatif dan positif dalam pemetaan topik, sekaligus mengembangkan visualisasi interaktif untuk menyoroti hubungan antar topik. Sebagai alat bantu, KNIME Analytics Platform digunakan untuk membandingkan metode ini dengan pendekatan lainnya, memberikan hasil yang lebih relevan dan terarah bagi rekomendasi konten.

## METODE PENELITIAN

### Topik Kontras berbasis Lift

Dalam penelitian ini, asosiasi antar topik, yang diwakili kata kunci / *keyword*, dari tiap dokumen dibangun menggunakan tiga metrik utama: Support, Confidence, dan Lift.

Diberikan dataset  $D$  berisi  $n$  dokumen,  $T$  adalah himpunan semua keyword unik di  $D$ , dan  $A$  dan  $B$  adalah keyword di  $T$  [9]:

- Support suatu keyword  $A$  ditulis  $\text{sup}(A)$  adalah banyaknya dokumen  $d_i \in D$  yang mengandung  $A$ . Support dihitung dengan persamaan (1)

$$\text{sup}(A) = \frac{|d_i|}{n}, A \subset d_i \in D \quad (1)$$

- Confidence  $A \rightarrow B$  adalah peluang jika keyword  $A$  muncul, maka  $B$  juga muncul, dengan rumus pada persamaan (2).

$$\text{conf}(A \rightarrow B) = \frac{\text{sup}(AB)}{\text{sup}(A)} \quad (2)$$

- Lift  $AB$ , adalah korelasi atau tingkat ketergantungan  $A$  dengan  $B$ , dihitung dengan rumus pada persamaan (3)

$$\text{lift}(AB) = \frac{\text{conf}(A \rightarrow B)}{\text{sup}(B)} \quad (3)$$

Dalam praktiknya, keyword  $A$ ,  $B$  dan  $AB$  secara umum disebut itemset, atau himpunan item. Secara umum, sebuah rule  $R$  ditulis dalam hubungan  $R: A \rightarrow B$ . Itemset  $A$  di sebelah kiri tanda panah disebut Anteseden (antecedent) atau bagian “Jika”, sedangkan itemset  $B$ , di sebelah kanan panah, disebut konsekuan (consequent) atau bagian “Maka”.

Metrik lift memberikan wawasan lebih lanjut tentang hubungan antara dua topik, terutama untuk menentukan apakah ada korelasi positif atau negatif. Lift mengukur seberapa banyak kemunculan  $A$  meningkatkan probabilitas kemunculan  $B$ , dibandingkan dengan jika  $A$  dan  $B$  terjadi secara independen.

Interpretasi dari nilai lift adalah sebagai berikut:

- *Lift* = 1:  
Tidak ada hubungan antara  $A$  dan  $B$ . Mereka terjadi secara independen.
- *Lift* > 1:  
Menunjukkan bahwa topik  $A$  dan  $B$  memiliki korelasi positif, artinya mereka sering muncul bersama-sama lebih sering daripada yang diharapkan jika mereka independen. Dalam konteks rekomendasi berita, hal ini berarti berita tentang  $B$  bisa ditawarkan ketika orang mencari tentang  $A$ . Misalnya, jika lift antara topik "Data Mining" ( $A$ ) dan "Machine Learning" ( $B$ ) lebih besar dari 1, ketika seseorang membaca berita tentang "Data Mining", sistem rekomendasi bisa menyarankan berita tentang "Machine Learning" karena ada kecenderungan kuat bahwa pembaca yang tertarik pada satu topik juga akan tertarik pada topik lainnya. Strategi ini berguna untuk memperkuat engagement dengan memberikan rekomendasi yang relevan dan meningkatkan peluang pembaca untuk tetap terlibat dengan konten.
- *Lift* < 1:  
Ini berarti  $A$  dan  $B$  jarang muncul bersama atau memiliki korelasi negatif. Dalam konteks rekomendasi, ini bisa diinterpretasikan bahwa kemunculan *berita tentang A justru mengurangi perhatian pada berita B*. Sebagai contoh, jika "Data Mining" ( $A$ ) dan "Data

Science" ( $B$ ) memiliki lift kurang dari 1, maka ketika pembaca membaca berita tentang "Data Mining", mereka mungkin kurang tertarik atau bahkan cenderung menghindari berita tentang "Data Science". Ini mengindikasikan bahwa berita tentang  $B$  (yang memiliki lift < 1 terhadap  $A$ ) tidak sebaiknya ditawarkan sebagai rekomendasi langsung setelah  $A$ . Sebaliknya, sistem rekomendasi harus mencari topik yang lebih relevan dengan  $A$  atau menawarkan berita yang kontras secara strategis untuk menarik perhatian pembaca.

Dalam pengembangan sistem rekomendasi konten, lift < 1, yang menunjukkan korelasi negatif, justru membuka peluang untuk menampilkan konten yang kontras atau berbeda. Dalam konteks konten berita, strategi ini efektif dalam menampilkan berita yang mungkin berbeda secara signifikan dengan yang sedang dibaca oleh pengguna. Pendekatan ini dapat memicu rasa penasaran pengguna, terutama dengan menawarkan berita yang bertolak belakang. Misalnya, ketika seorang pembaca sedang membaca artikel pro tentang "Penggunaan AI di Industri," berita yang membahas risiko atau masalah etika AI dapat ditawarkan sebagai sudut pandang alternatif. Hal ini memberikan novelty yang signifikan dan dapat menarik perhatian pembaca yang tertarik mengeksplorasi perspektif lain.

Selain itu, ketika lift > 1, di mana topik yang sama sering muncul bersama, memberikan rekomendasi berita yang serupa dapat menyebabkan kebosanan bagi pembaca. Oleh karena itu, menyajikan berita yang berbeda atau mengejutkan dapat membantu menghindari repetisi topik dan mencegah apa yang disebut "saturation effect." Dalam kasus ini, pembaca merasa sudah mendapatkan cukup informasi tentang topik tertentu, sehingga variasi konten dapat meningkatkan minat mereka.

Sistem rekomendasi juga dapat menerapkan strategi kontrasif, di mana topik-topik kontras ditawarkan secara selektif setelah pengguna membaca beberapa artikel tentang satu topik tertentu. Dengan cara ini, pembaca dapat

mengeksplorasi topik-topik yang lebih beragam dan memperluas wawasan mereka melalui konten yang disajikan secara strategis.

### Rancangan Eksperimen

Eksperimen dilakukan menggunakan KNIME Analytics Platform untuk membandingkan performa LDA dengan metode berbasis AR yang diusulkan. Dataset yang digunakan terdiri dari dua jenis:

1. Dataset 1 Hasil Pencarian Google: berisi 25 hasil pencarian Google dengan kata kunci "Data Mining (DM) versus Data Science (DS)" dalam bahasa Inggris. Hasil pencarian ini mencakup berbagai artikel, blog, dan referensi yang terkait dengan kedua topik tersebut. Data ini dipilih untuk dilihat bagaimana kedua istilah ini dibandingkan dan diinterpretasikan dalam berbagai sumber online. Setiap hasil pencarian diproses sebagai satu dokumen.
2. Dataset 2 Judul Berita Bahasa Indonesia bertema Politik Nasional 2024, yang diperoleh dari sebuah media online di Kota Malang. Judul berita digunakan sebagai representasi isi berita, karena dalam praktik penulisan, judul sering kali dibuat dengan kata-kata yang unik dan dianggap mewakili keseluruhan isi berita. Dataset terdiri atas 780 baris berita.

### Workflow KNIME

Pengujian atas dua metode yang dibandingkan, dilakukan menggunakan KNIME dengan cara membuat workflow (alur kerja), yang dijelaskan dalam Gambar di Appendix. Workflow KNIME berfungsi sebagai rangkaian alur yang menghubungkan berbagai komponen pemrosesan data secara visual.

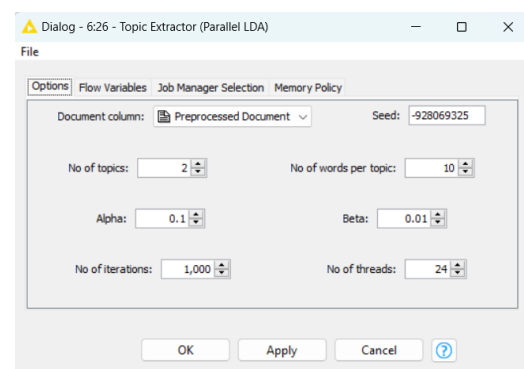
*Workflow* dibagi empat bagian: Pra pemrosesan, Metode LDA dan Metode yang diusulkan.

1. **Pra-Pemrosesan:** node yang digunakan antara lain berfungsi untuk membersihkan tiap baris data dari tanda baca, angka, dan spasi kosong; menghilangkan stop words; dan

tokenisasi, yaitu memecah kata ke dalam bentuk *Set* agar tiap kata tersimpan secara unik.

2. **Metode LDA:** node yang digunakan adalah *Topic Extractor* (Parallel LDA), yaitu node bawaan KNIME untuk mengekstrak topik (*keyword*) dari data tekstual. Luaran dari node ini, divisualisasikan dengan node Scatter plot.

Pengaturan yang diberikan pada Node Parallel LDA, sebagaimana dalam Gambar 1, antara lain: Document column: berisi nama kolom yang berisi dokumen yang telah dipra-proses; No of topics: jumlah topik yang diinginkan; No of words per topic: jumlah kata kunci yang ingin diekstrak untuk setiap topik dan sekaligus merepresentasikan topik; Alpha: menentukan distribusi topik di setiap dokumen. Nilai alpha yang lebih kecil (misalnya 0.1) cenderung menghasilkan distribusi topik yang lebih jarang dalam setiap dokumen, artinya sebuah dokumen lebih mungkin terkait dengan hanya sedikit topik. Beta: prior distribusi kata dalam setiap topik. Beta yang lebih kecil (misalnya 0.001) mendorong distribusi kata yang lebih jarang dalam topik; No of threads: membagi kumpulan dokumen menjadi beberapa thread untuk pemrosesan paralel, mempercepat waktu komputasi.



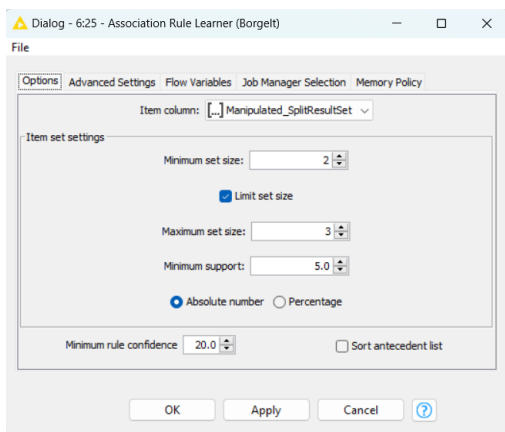
**Gambar 1.** Pengaturan Parameter di Node *Topic Extractor* (Parallel LDA)

3. **Metode yang Diusulkan:** terdiri dari rangkaian Node, termasuk Node AR Learner (*Borgelt*), dimana daftar *rule* yang dihasilkan kemudian dibagi

menjadi dua jalur pengamatan menggunakan Node Row Filter. Untuk memvisualisasikan hubungan antar *keyword*, digunakan plot koordinat paralel yang dihasilkan oleh Node Paralel Coordinate Plot.

4. **Pasca pemrosesan dan visualisasi:** digunakan pemberian warna terhadap hasil dari tiap metode, dengan tujuan agar visualisasi lebih efektif untuk diinterpretasikan. Pemberian warna hanya dapat dilakukan pada maksimal 60 keyword saja.

Pengaturan yang diberikan pada Node AR Learner (Borgelt) diberikan pada Gambar 2, dengan penjelasan ringkas sebagai berikut: Items column, adalah kolom berisi data yang akan dipelajari AR-nya. Kolom ini adalah hasil pra-proses, dan berbentuk Set dari keyword, atau dalam konsep data mining, sering disebut Bag of words. Minimum dan maximum set size, adalah jumlah keyword minimal dan maksimal di dalam itemset (himpunan keyword) yang terbentuk; minimum support dan minimum confidence. Node ini akan menghasilkan tabel rules, yang dilengkapi dengan nilai support, confidence dan lift.



**Gambar 2.** Pengaturan di Node Association Rule Learner (Borgelt)

### HASIL DAN PEMBAHASAN Pra Pemrosesan Data

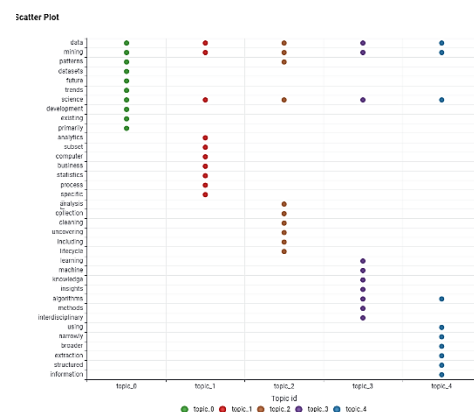
Rules yang dihasilkan oleh Node Association Rule Learning (Borgelt) adalah sebanyak 2572 rules dari Dataset ke-1, dan 6192 rules dari Dataset ke-2, sesuai

pengaturan dalam Gambar 1 dan 2. Namun pada praktiknya, untuk tujuan pemvisualisasi, jumlah rules ini mengalami penyaringan di Node Row Filter, hingga hanya maksimal 60 keyword saja.

### Dataset ke-1: DM vs DS

Dataset ke-1 berisi hasil pencarian di Google menggunakan kata kunci “Data mining versus Data science”. Hasil pemetaan topik terhadap dataset ini, menggunakan LDA, diberikan di Gambar 3; setelah itu, *keyword* yang terkandung dalam tiap topik yang terbentuk dianalisis keterhubungannya.

Misalnya pada topik 0, memiliki keywords: *data, mining, patterns, datasets, future, trends, science, development, existing, primarily*. Analisis secara ringkas menyimpulkan bahwa topik ini menekankan pada identifikasi pola dalam data melalui teknik data mining, dengan fokus pada tren masa depan dan pengembangan ilmu pengetahuan. Kata-kata seperti "future" dan "trends" menunjukkan adanya perhatian terhadap perkembangan teknologi dan bagaimana pola dari dataset yang ada dapat memberikan wawasan tentang inovasi di masa depan. Topik ini menggambarkan bagaimana sains data terus berkembang dengan eksplorasi dataset yang sudah ada.



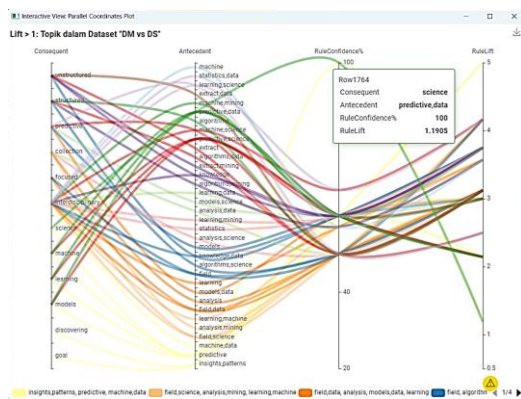
**Gambar 3.** Topic mapping dengan LDA untuk tema DM vs DS

Topik 1 dengan keywords: *data, mining, analytics, science, subset, computer, business, statistics, process, dan specific*, diinterpretasikan bahwa topik 1 memiliki kaitan dengan penerapan data mining dan analitik dalam konteks bisnis. Fokus pada

kata-kata seperti "business" dan "statistics" menunjukkan bahwa topik ini mengacu pada bagaimana proses statistik dan komputasi digunakan untuk menganalisis data bisnis dan menghasilkan keputusan berbasis data.

Tampilan berbeda disajikan untuk hasil pemetaan topik dengan metode AR learning, seperti pada Gambar 4 dan 5 yang memetakan rules dengan Lift > 1. Di sini tidak diperlihatkan klasifikasi kata ke dalam topik ke-1, 2 dan seterusnya seperti dalam LDA; namun hubungan antara anteseden (keyword X) dan konsekuen (Y) dari rules dalam nilai Confidence dan Lift. Nilai Confidence menunjukkan peluang bahwa keyword Y akan muncul, jika keyword X muncul.

Misalnya dalam Gambar 4, jika topik yang mengandung keyword [data, predictive] dibahas, maka 100% topik ini juga berkaitan dengan keyword [science]. Topik yang relevan dengan ketiga kata ini adalah *predictive data science*. Lift > 1 juga menunjukkan bahwa keyword "data, predictive" akan selalu muncul dengan "science".



**Gambar 4.** Topic mapping dengan AR Learning (Lift > 1) untuk tema DM vs DS

Contoh lain, di Gambar 5, didemonstrasikan bahwa keyword "interdisciplinary" berpeluang muncul sebesar 66.7% dengan "machine, science", dengan Lift = 4.1667. Topik yang relevan dengan keyword ini adalah bahwa *Data science is an interdisciplinary field*.

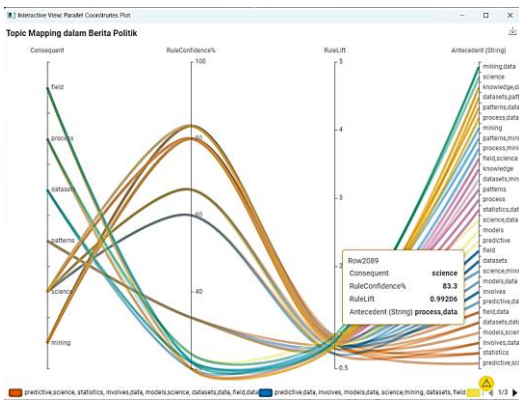


**Gambar 5.** Hubungan antara keyword "machine, science" dan "interdisciplinary"

Dari pola-pola ini, dapat disimpulkan bahwa AR learning tidak hanya membantu mengidentifikasi keterkaitan antar kata atau konsep yang sering muncul bersama, tetapi juga menyoroti bagaimana konsep-konsep ini saling mendukung dalam konteks yang lebih luas. Misalnya, hubungan antara structured dan unstructured data, serta hubungan machine learning dengan predictive dan interdisciplinary approaches. Pola ini tidak dapat diungkapkan oleh metode seperti LDA, yang hanya berfokus pada distribusi kata tanpa memperhitungkan hubungan antara mereka.

Contoh hasil pemetaan topik berbasis rules dengan nilai Lift < 1 diberikan pada Gambar 6. Di sini diperlihatkan hubungan negatif antara keyword "data, process" dan "science" dengan Lift = 0.99206, yang bermakna hubungan antara "data, process" dan "science" menunjukkan korelasi negatif yang sangat lemah, yang berarti kemunculan "data" dan "process" sedikit mengurangi kemungkinan munculnya "science" dalam topik yang sama. Meskipun Lift mendekati 1, hal ini menunjukkan bahwa kedua set kata ini tidak terlalu sering muncul bersama dibandingkan dengan apa yang diharapkan secara statistik.

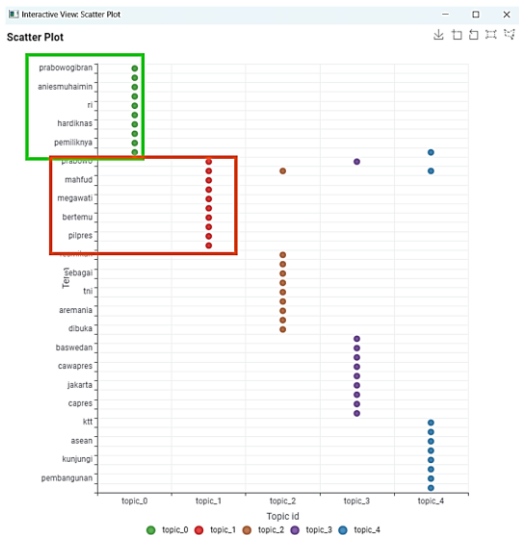
Namun, dengan confidence sebesar 80%, ini menunjukkan bahwa meskipun Lift = 0.99206 menandakan hubungan negatif yang lemah antara "data, process" dan "science", ada kemungkinan 80% bahwa ketika "data" dan "process" muncul, "science" tetap akan muncul bersama mereka. Artinya, meskipun asosiasi antara ketiga kata ini lebih jarang terjadi dibandingkan ekspektasi, ketika "data" dan "process" dibahas, ada kemungkinan



**Gambar 6. Pemetaan Topik berbasis AR dengan Rules bernilai Lift < 1**

**Dataset ke-2: Berita Politik 2024**

Metode LDA memetakan topik seputar berita politik 2024 ke dalam lima topik, seperti dalam Gambar 7.



**Gambar 7. Topic mapping dengan LDA untuk Berita Politik 2024**

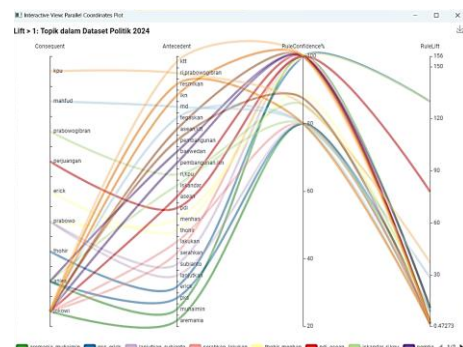
Interpretasi terhadap beberapa topik misalnya sebagai berikut:

- Topik 0, keywords: prabowogibran, kpu, aniesmuhaimin, bersama, ri, ppp, hardiknas, asyiknya, pemiliknya, dan pranowo. Topik ini tampaknya berfokus pada konstelasi politik terbaru terkait dengan pasangan Prabowo-Gibran dan Anies-Muhaimin.
- Topik 1, keywords: prabowo, jokowi, mahfud, hoaks, megawati, ganjarmahfud, bertemu, md, pilpres, yogyakarta. Topik ini mengarah pada pertemuan politik antar tokoh-tokoh

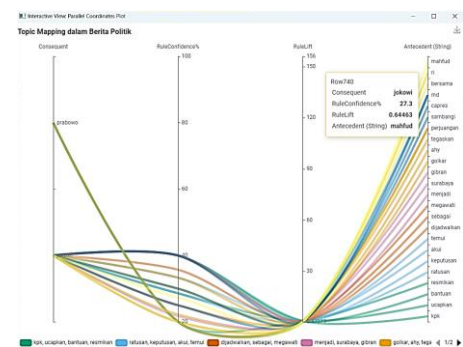
besar seperti Prabowo, Jokowi, Mahfud MD, dan Megawati.

Hasil pemetaan topik menggunakan AR learning dengan nilai Lift > 1, diberikan dalam Gambar 8. Interpretasi topik-topik yang terkandung dalam peta ini antara lain diberikan sebagai berikut:

- Keyword anteseden “muhaimin” dan konsekuen “anies” muncul dengan confidence 100%, dan Lift = 11.143 menunjukkan topik mengenai koalisi dan Pasangan Capres-Cawapres ini; dengan kata lain, saat muncul berita mengenai Muhaimin, maka berita Anies juga muncul.
- Konsekuen “Jokowi” berkaitan dengan beberapa anteseden berbeda, seperti [“pembangunan”, “ikn”], [“asean”, “ktt”], menunjukkan bahwa ketika berita mengenai ‘pembangunan ikn’ atau ‘ktt asean’ muncul, maka berita mengenai Presiden Joko Widodo juga muncul, dengan confidence berkisar antara 80-100%, dan Lift > 1.



**Gambar 8. Topic mapping dengan AR Learning (Lift > 1) untuk Berita Politik 2024**



**Gambar 9. Topic mapping dengan Lift < 1 untuk Berita Politik 2024**

Di sisi lain, jika berita di atas adalah berita *mainstream*, maka beberapa pembaca mungkin akan tertarik atau penasaran dengan berita yang bertolak belakang, yang dapat diperoleh dengan menyelidiki  $Lift < 1$ . Misalnya berita tentang “Mahfud M.D” yang jika muncul, maka berita mengenai “Jokowi” tidak muncul dengan  $Lift$  hanya sebesar 0.64463, dan confidence sebesar hanya 27.3%. Bagi pembaca yang mungkin sudah familiar dengan berita *mainstream* tentang Presiden Jokowi, hubungan negatif ini dapat menarik perhatian mereka untuk mencari perspektif alternatif atau berita kontras yang melibatkan Mahfud M.D. Peta topik dengan Rule bernilai  $Lift < 1$  diberikan dalam Gambar 9.

Dalam pengelolaan sistem rekomendasi konten di portal berita daring, pengelola konten dapat menggunakan pemetaan topik sebagai strategi utama untuk mengelola pengalaman pengguna. Dengan terlebih dahulu memetakan topik di dalam konten berita yang tersedia, pengelola dapat mengidentifikasi topik *mainstream* yang sering dibaca oleh mayoritas pengguna. Setelah itu, mereka dapat menyesuaikan laman portal berita dengan memberikan rekomendasi yang lebih *mainstream* bagi pembaca yang menyukai topik populer atau memberikan rekomendasi yang kontras untuk pembaca yang mungkin tertarik dengan sudut pandang berbeda atau alternatif dari berita arus utama.

## PENUTUP

Metode AR *Learning* dengan  $Lift$  positif dan negatif memberikan keunggulan signifikan dibandingkan dengan LDA dalam memahami hubungan antar topik. AR *learning* tidak hanya mampu mengidentifikasi asosiasi yang sering muncul bersama, tetapi juga mengungkapkan kontras dan hubungan negatif yang sering tidak terdeteksi oleh LDA. Pendekatan ini memungkinkan pengelola konten di media online untuk lebih fleksibel dalam merekomendasikan konten *mainstream* yang populer atau konten kontras yang menawarkan perspektif alternatif kepada pembaca, sehingga meningkatkan keterlibatan dan kepuasan pengguna.

## DAFTAR PUSTAKA

- [1] G. Nanda, A. Jaiswal, H. Castellanos, Y. Zhou, A. Choi, and A. J. Magana, “Evaluating the Coverage and Depth of Latent Dirichlet Allocation Topic Model in Comparison with Human Coding of Qualitative Data: The Case of Education Research,” *Mach. Learn. Knowl. Extr.*, vol. 5, no. 2, 2023, doi: 10.3390/make5020029.
- [2] D. M. Blei, “Introduction to Probabilistic Topic Models,” *Commun. ACM*, vol. 55, no. 4, 2012.
- [3] H. Jelodar *et al.*, “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimed. Tools Appl.*, vol. 78, no. 11, 2019, doi: 10.1007/s11042-018-6894-4.
- [4] Q. Xie, J. Huang, T. Saha, and S. Ananiadou, “GRETEL: Graph Contrastive Topic Enhanced Language Model for Long Document Extractive Summarization,” in *Proceedings - International Conference on Computational Linguistics, COLING, 2022*.
- [5] X. Wu, T. Nguyen, and A. T. Luu, “A survey on neural topic models: methods, applications, and challenges,” *Artif. Intell. Rev.*, vol. 57, no. 2, 2024, doi: 10.1007/s10462-023-10661-7.
- [6] J. Al Qundus, S. Peikert, and A. Paschke, “AI supported Topic Modeling using KNIME-Workflows,” in *CEUR Workshop Proceedings, CEUR-WS, 2020*.
- [7] Y. Gao, Y. Li, R. Y. K. Lau, Y. Xu, and M. A. Bashar, “Finding semantically valid and relevant topics by Association-based Topic Selection model,” *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, 2017, doi: 10.1145/3094786.
- [8] L. D. Adistia, T. M. Akhriza, and S. Jatmiko, “Sistem Rekomendasi Buku untuk Perpustakaan Perguruan Tinggi Berbasis Association Rule,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, 2019, doi: 10.29207/resti.v3i2.971.



- [9] M. Husni, T. M. Akhriza, S. Madenda, and E. P. Wibowo, "Improving Recentness of the ICT Book Recommendation using an Adaptive Rules-based Recommender System," *Int. J. Comput. Appl. Technol.*, vol. (In Press), 2022.
- [10] T. M. Akhriza and I. D. Mumpuni, "A Time-Window Approach to Recommending Emerging and On-the-rise Items," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022.
- [11] T. A. Armanda, I. P. Wardhani, T. M. Akhriza, and T. M. A. Admira, "Recurrent Session Approach to Generative Association Rule based Recommendation," *Knowl. Eng. Data Sci.*, vol. 6, no. 2, pp. 199–214, 2023, [Online]. Available: <http://repo.stimata.ac.id/id/eprint/376/>