PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

Analisis Pemodelan Topik Mengenai Negara Dengan Kecerdasan Buatan Terbaik Menggunakan *Latent Dirichlet Allocation* (LDA) di *Orange Data Mining*

I Made Mustika Kerta Astawa[1], Mega Suci Lestari[2], Efza Nur Agatya[3]
[1,2,3]Program Studi Teknik Informatika S-2, Universitas Pamulang
e-mail: made.mustika19@gmail.com[1], mega.sucilestari@gmail.com[2], efzanuragastya@gmail.com[3]

***Abstrak*—Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan topik-topik utama dalam dokumen teks yang membahas negara-negara dengan penerapan kecerdasan artifisial (AI) terbaik. Dengan menggunakan pendekatan *Latent Dirichlet Allocation* (LDA) pada platform Orange Data Mining, data dikumpulkan dari 11 artikel daring yang relevan. Proses *preprocessing* dilakukan untuk membersihkan teks sebelum diubah menjadi representasi numerik menggunakan TF-IDF. Hasil analisis berhasil mengidentifikasi 10 topik utama, yang mencakup isu seperti pendidikan AI, peringkat universitas, kesiapan nasional, serta peluang kerja di bidang AI. Evaluasi model menunjukkan koherensi topik yang baik. Penelitian ini membuktikan bahwa pendekatan LDA dengan Orange efektif dalam mengeksplorasi informasi tematik tersembunyi dalam kumpulan dokumen teks AI..**

***Kata Kunci*— Kecerdasan Artifisial, Topik Modeling, *Latent Dirichlet Allocation* (LDA), Analisis Teks, Orange Data Mining.**

I. PENDAHULUAN

Di era Revolusi Industri 4.0, kecerdasan artifisial (*Artificial Intelligence*/AI) menjadi salah satu pilar utama dalam mendorong transformasi digital di berbagai sektor, mulai dari pemerintahan, industri, kesehatan, pendidikan, hingga pertahanan. AI tidak hanya dipandang sebagai alat bantu otomatisasi, tetapi juga sebagai teknologi strategis yang menentukan daya saing sebuah negara di kancah global. Oleh karena itu, banyak negara berlomba-lomba mengembangkan dan menerapkan AI sebagai bagian dari strategi nasional mereka.

Negara-negara seperti Amerika Serikat, China, Inggris, Jerman, dan Kanada telah menginvestasikan sumber daya yang sangat besar dalam hal riset, infrastruktur, pendidikan, dan kebijakan untuk mengakselerasi pengembangan AI. Amerika Serikat unggul dalam inovasi berbasis industri dan start-up teknologi, sementara China mengedepankan kebijakan negara dan integrasi teknologi AI dalam skala nasional. Negara-negara Eropa juga tidak ketinggalan, dengan menekankan pada etika AI dan keberlanjutan teknologi.

Dengan berkembangnya literatur dan publikasi mengenai AI di berbagai negara, muncul kebutuhan untuk mengekstraksi informasi secara efisien dari dokumen-dokumen tersebut. Salah satu pendekatan yang dapat digunakan adalah topik modeling, sebuah metode dalam Text Mining yang bertujuan untuk mengidentifikasi tema atau topik tersembunyi yang terkandung dalam dokumen teks tanpa perlu membaca setiap dokumen secara manual. Topik modeling membantu menyederhanakan data dalam jumlah besar dan memperjelas struktur tematik yang tersembunyi.

Salah satu metode populer dalam topik modeling adalah *Latent Dirichlet Allocation* (LDA). LDA merupakan model probabilistik generatif yang mengasumsikan bahwa setiap dokumen merupakan campuran dari beberapa topik, dan setiap topik merupakan distribusi probabilistik atas kata-kata. LDA banyak digunakan dalam analisis dokumen seperti berita, publikasi ilmiah, media sosial, dan laporan kebijakan.

Dalam konteks ini, penelitian dilakukan dengan menggunakan Orange Data Mining, sebuah platform visualisasi dan analisis data yang memungkinkan pengguna untuk melakukan analisis teks, termasuk topik modeling, secara intuitif tanpa perlu coding. Orange menyediakan berbagai widget untuk memproses data teks, melakukan preprocessing, membentuk representasi Bag of Words atau TF-IDF, dan menerapkan model LDA.

Dengan menggunakan Orange dan metode LDA, analisis ini bertujuan untuk mengidentifikasi topik-topik utama dalam dokumen yang membahas negara-negara dengan AI terbaik. Melalui hasil yang diperoleh, kita dapat melihat kecenderungan

tematik yang berkembang di masing-masing negara, seperti fokus pada inovasi teknologi, kebijakan pemerintah, riset, pendidikan, atau pengembangan sumber daya manusia.

Dengan kata lain, penelitian ini tidak hanya memberikan gambaran tentang struktur tematik dokumen teks, tetapi juga membuka peluang untuk memahami prioritas dan pendekatan strategis negara-negara tersebut dalam menghadapi era kecerdasan buatan.

II. METODE PENELITIAN

1. **Pengumpulan Data**
   Mengumpulkan data teks yang relevan terkait topik negara dan kecerdasan buatan terbaik. Data dapat diperoleh dari sumber online seperti artikel berita, jurnal, media sosial (misalnya Twitter), atau dokumen digital lainnya.
2. **Preprocessing Data**
   Melakukan pembersihan data teks dengan tahapan:
   a. Menghapus tanda baca, angka, dan karakter khusus
   b. Menghilangkan kata-kata umum (stopwords)
   c. Melakukan tokenisasi (memecah teks menjadi kata-kata)
   d. Stemming atau lemmatization untuk mengubah kata ke bentuk dasar
   e. Filtering kata yang sangat sering atau sangat jarang muncul agar model fokus pada kata relevan
3. **Import Data ke Orange Data Mining**
   Memasukkan data hasil preprocessing ke dalam Orange Data Mining menggunakan widget *File* atau *Corpus*.
4. **Pemodelan Topik dengan LDA**
   Menggunakan widget *Topic Modelling* di Orange untuk menerapkan algoritma *Latent Dirichlet Allocation* (LDA).
   a. Menentukan jumlah topik yang diinginkan
   b. Mengatur parameter seperti iterasi dan *hyperparameter* (alpha, beta)
   c. Menjalankan proses pemodelan untuk mengelompokkan kata-kata ke dalam topik-topik laten
5. **Visualisasi dan Interpretasi Hasil**
   Menggunakan widget visualisasi seperti *Word Cloud*, *Topic Distribution*, dan *LDAvis* untuk melihat kata-kata dominan dalam tiap topik dan distribusi topik dalam dokumen. Interpretasi dilakukan untuk memahami tema utama yang muncul.
6. **Evaluasi Model**
   Mengevaluasi kualitas model dengan mengukur koherensi topik dan relevansi kata-kata dalam setiap topik. Jika perlu, melakukan tuning parameter untuk meningkatkan hasil.

III. PEMBAHASAN

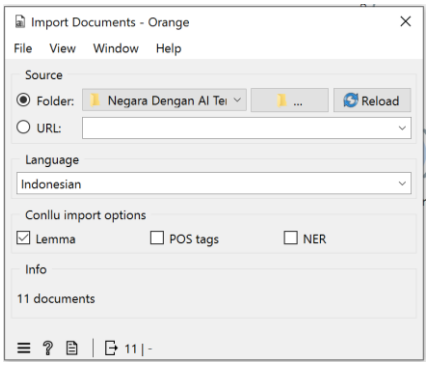
A. Pengumpulan *Dataset*

Negara Dengan Kecerdasan terbaik Terbaik pada Link sebagai berikut :

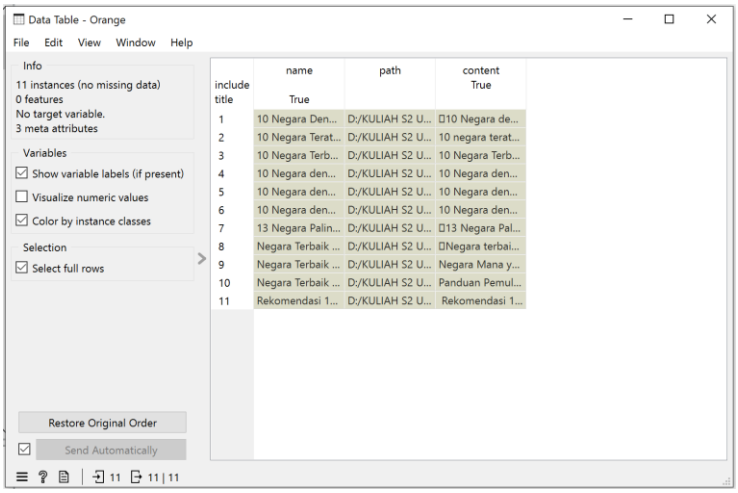
1. https://www.detik.com/edu/detikpedia/d-7955860/10-negara-dengan-skill-ai-terbaik-di-asia-versi-coursera-ada-indonesia#google_vignette
2. https://rejolut-com.translate.goog/blog/13-top-ai-countries/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
3. https://www-siecindia-com.translate.goog/blogs/which-country-is-best-to-study-artificial-intelligence-and-machine-learning?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
4. https://www.floreseditorial.com/news/39715325453/10-negara-dengan-skill-ai-terbaik-di-asia-versi-coursera-2025-singapura-teratas-indonesia-mulai-unjuk-gigi-di-peringkat-11
5. https://metrokompas.id/rekomendasi-10-negara-asia-dengan-keahlian-ai-terbaik-ada-indonesia/
6. https://herovired-com.translate.goog/learning-hub/blogs/beginners-guide-to-best-countries-for-artificial-intelligence-jobs/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc&_x_tr_hist=true
7. https://studyinternational-com.translate.goog/news/best-countries-to-study-ai/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
8. https://data.goodstats.id/statistic/10-negara-dengan-indeks-kesiapan-ai-terbaik-bo4xG
9. https://dataindonesia.id/tenaga-kerja/detail/deretan-negara-dengan-jumlah-ai-tech-talent-terbanyak-pada-2025
10. https://tekno.sindonews.com/read/1437521/207/10-negara-paling-maju-dalam-perlombaan-ai-israel-kalah-telak-dari-china-1723878531
11. https://www-admissify-com.translate.goog/blog/best-countries-to-study-artificial-intelligence/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

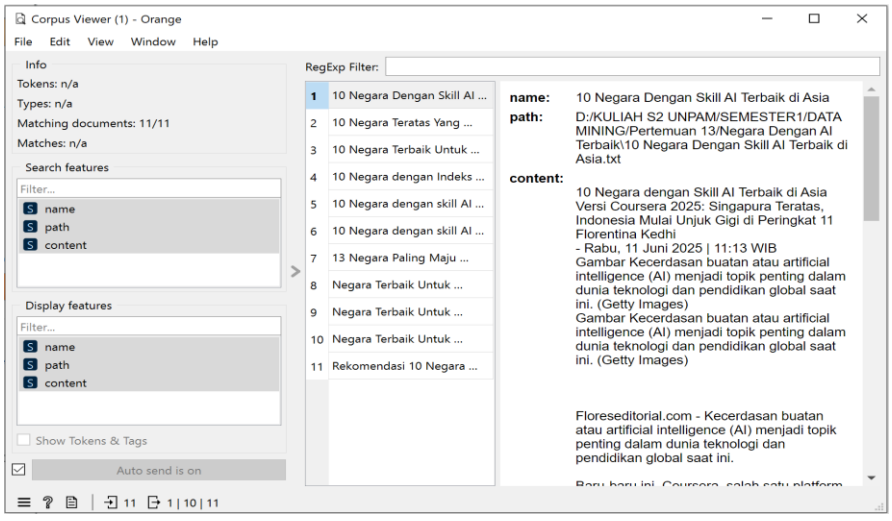
Dataset terdiri dari 11 Data teks dengan menyimpan teks berita dengan format .txt dalam 1 Folder Data.



Gambar 1. Import Dokumen



Gambar 2. Tabel Dataset



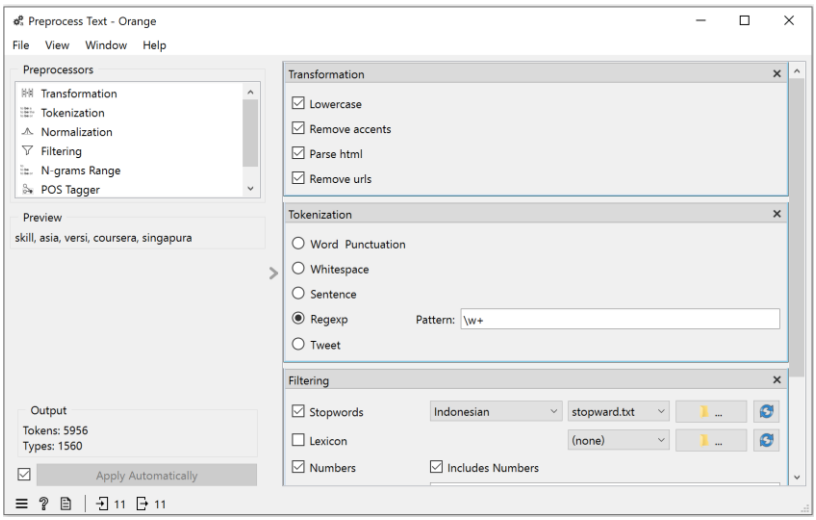
Gambar 3. Corpus Viewer

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

B. Preprocessing

Dilakukan preprocessing data teks dengan tahapan sebagai berikut:

a. Transformation (*Lowercase*, *Remove Accents*, *Parse html* dan *Remove urls*)
b. Tokenisasi
c. Stopwords removal
d. Normalisasi kata tidak baku
e. Stemming

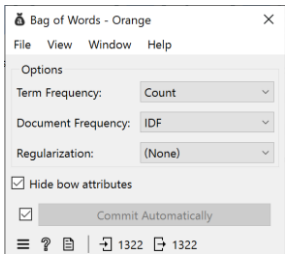


Gambar 4. Prepocessing Teks

.

C. Ekstraksi Fitur

Menggunakan metode **TF-IDF** untuk mengubah teks menjadi representasi numerik. TF-IDF mempertimbangkan pentingnya kata dalam dokumen relatif terhadap seluruh korpus.



Gambar 5. Bag of Word

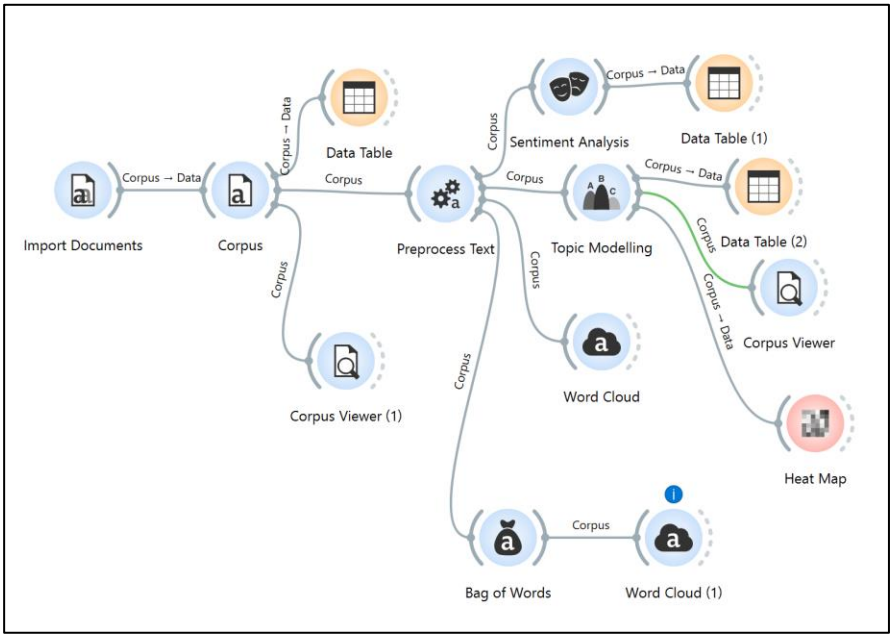
D. Pemodelan Orange Data Mining

Pada penelitian ini digunakan algoritma *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi dan mengekstrak topik atau tema yang tersembunyi (*latent topic*) dalam sebuah koleksi dokumen, teks maupun dataset. LDA (*Latent Dirichlet Allocation*) adalah salah satu metode topik modeling yang digunakan untuk menemukan struktur tersembunyi (topik) dalam kumpulan dokumen teks yang besar.

Manfaat LDA:

a. Menemukan pola atau tema tersembunyi dalam kumpulan teks besar.
b. Berguna dalam analisis dokumen, klasifikasi otomatis, rekomendasi, hingga eksplorasi isi teks.
c. Cocok digunakan dalam berbagai bidang: jurnalistik, penelitian akademik, media sosial, manajemen pengetahuan, dsb.

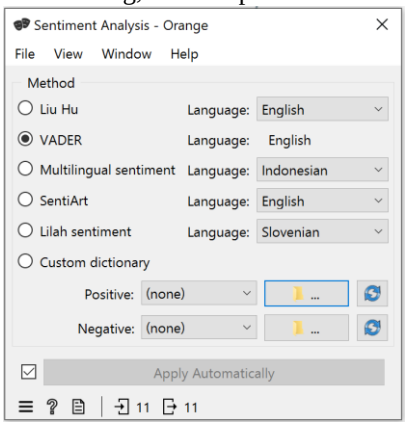
PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845



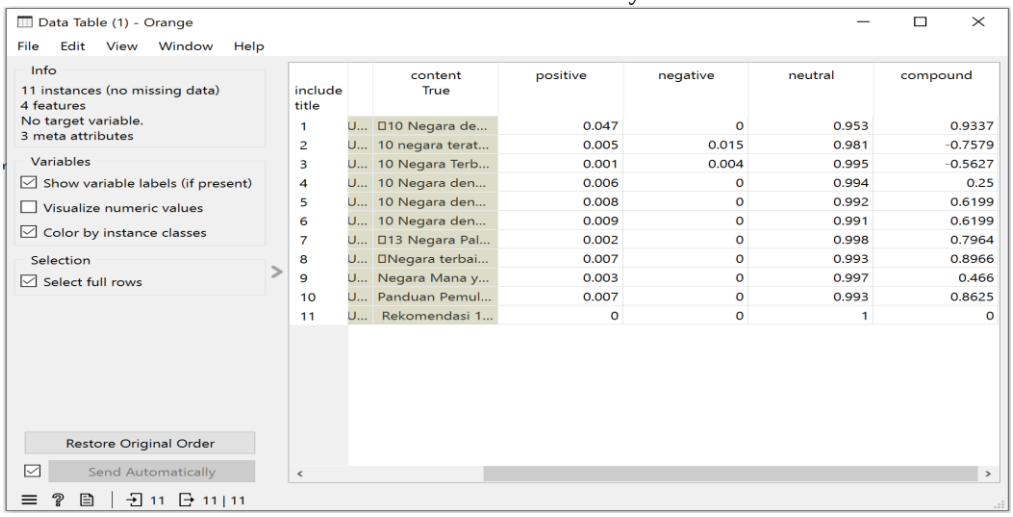
Gambar 6. Pemodelan Pada Orange

IV. HASIL DAN ANALISIS

Setelah dilakunan Pemodelan pada Orange Data Mining, maka dapat diketahui hasil dan analisa sebagai berikut :



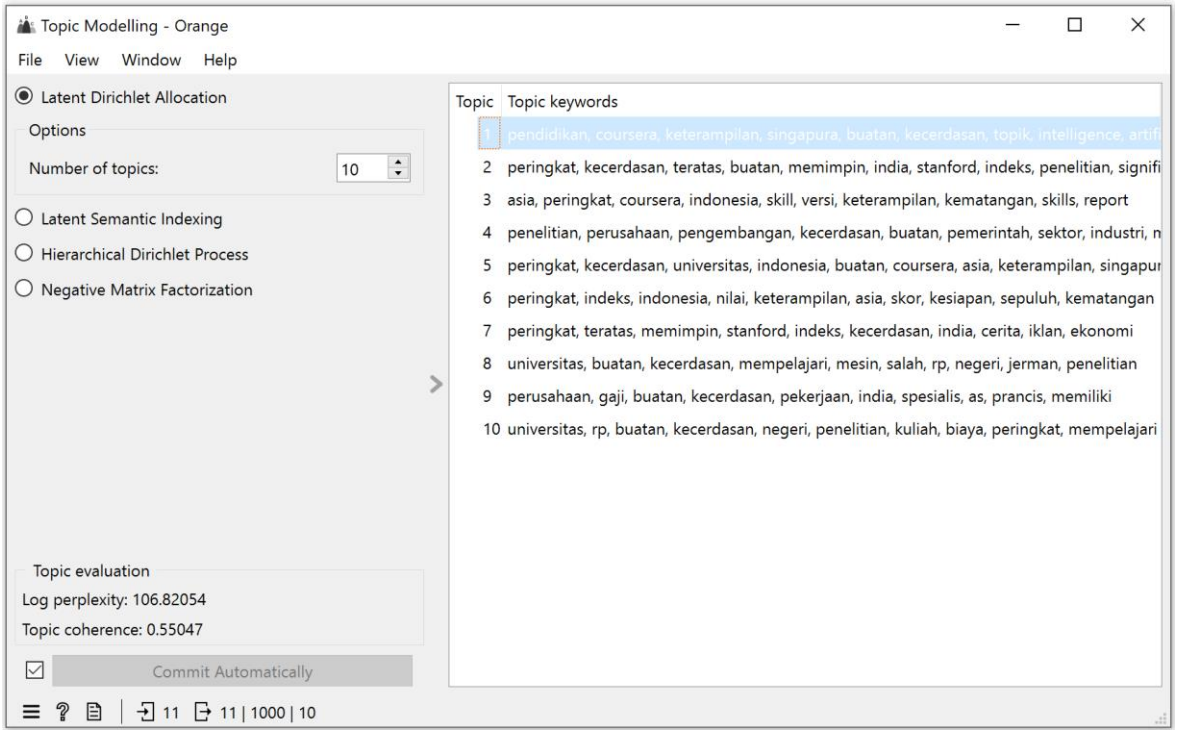
Gambar 7. Sentiment Analysis



Gambar 8. Tabel Sentiment Analysis

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

Tabel menunjukkan hasil analisis sentimen dengan 4 fitur utama per dokumen:
positive: tingkat sentimen positif (0–1)
Ada 1 dokumen dengan skor positif > 0.04 (dokumen pertama: 0.047), menunjukkan sedikit nada optimisme atau pujian.
negative: tingkat sentimen negatif (0–1)
Hanya dua dokumen yang mengandung sentimen negatif (0.015 dan 0.004), dengan skor sangat rendah. Ini menunjukkan bahwa nada negatif sangat jarang.
neutral: tingkat sentimen netral (0–1)
Semua dokumen memiliki nilai netral tinggi (≥ 0.95), menunjukkan bahwa isi cenderung bersifat informatif, deskriptif, atau objektif.
compound: skor gabungan (skala -1 sampai 1), semakin mendekati 1 berarti lebih positif
Compound berkisar dari -0.75 (cukup negatif) hingga 0.93 (sangat positif). Terdapat 2 artikel dengan compound negatif.



Gambar 9. Topic Modelling Dengan LDA

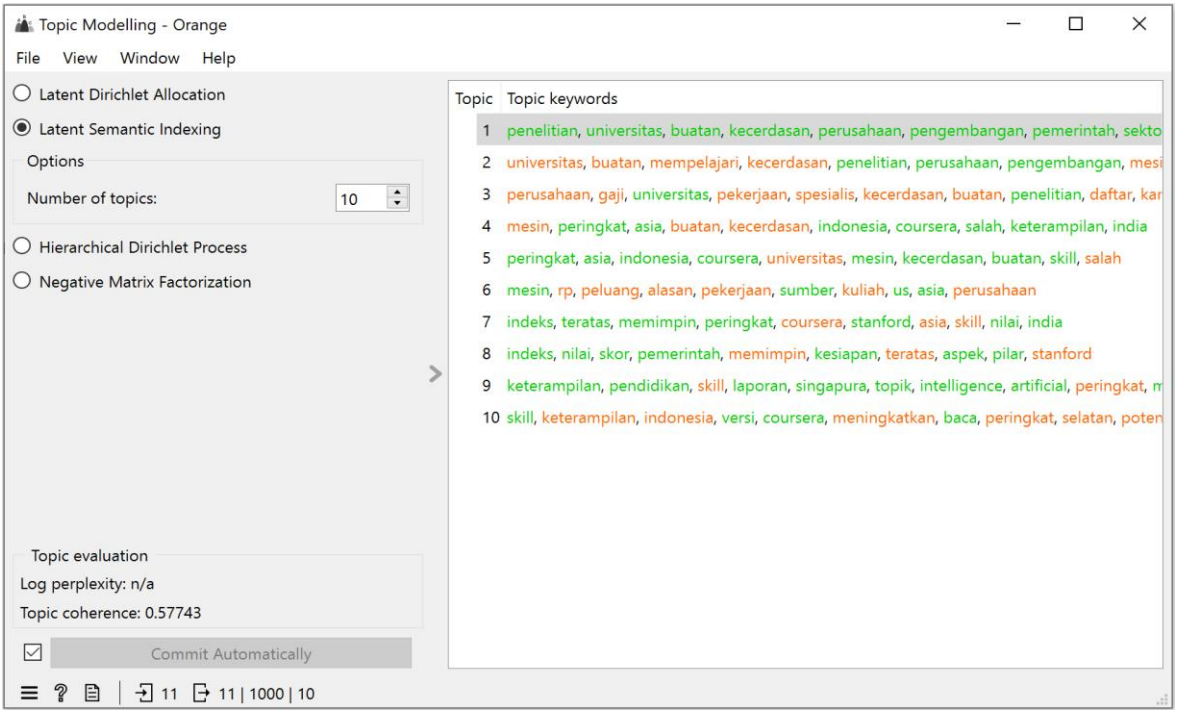
LDA berhasil mengekstrak 10 topik dari kumpulan dokumen dengan kata-kata kunci (keywords) utama yang mewakili masing-masing topik. Nilai evaluasi model:
Log Perplexity: 106.82054, Semakin kecil biasanya semakin baik (tapi relatif tergantung konteks).
Topic Coherence: 0.55047, Nilai koherensi yang cukup baik (nilai antara 0.4–0.6 sudah cukup informatif).

Analisis Per Topik :

No	Topik Utama (Keywords)	Makna/Interpretasi
1	pendidikan, coursera, keterampilan, singapura, buatan, kecerdasan, topik, intelligence	Topik edukasi dan pelatihan AI, khususnya berbasis platform online seperti Coursera. Singapura menonjol sebagai negara contoh.
2	peringkat, kecerdasan, teratas, buatan, memimpin, india, stanford	Topik peringkat negara/instansi AI terbaik, termasuk India dan Stanford. Fokus pada kompetisi global.
3	asia, coursera, indonesia, skill, versi, keterampilan	Fokus kawasan Asia dan platform pembelajaran keterampilan digital/AI, dengan Indonesia sebagai konteks.
4	penelitian, perusahaan,	Topik inovasi dan riset AI dalam industri dan

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

No	Topik Utama (Keywords)	Makna/Interpretasi
	pengembangan, pemerintah, sektor, industri	kebijakan pemerintah, seperti peran sektor privat dan publik.
5	peringkat, universitas, indonesia, buatan, coursera	Pemeringkatan universitas dalam bidang AI di Indonesia, serta penggunaan platform seperti Coursera.
6	indeks, nilai, skor, kesiapan, sepuluh, kematangan	Topik indeks kesiapan AI nasional, menilai seberapa siap negara dalam mengadopsi AI secara strategis.
7	peringkat, teratas, stanford, cerita, iklan, ekonomi	Topik media dan narasi kesuksesan AI, dengan keterkaitan ke kampus besar seperti Stanford dan dampak ekonomi.
8	universitas, buatan, mempelajari, mesin, salah, negeri	Pembelajaran mesin dan AI di tingkat universitas, baik lokal maupun internasional.
9	perusahaan, gaji, pekerjaan, spesialis, prancis	Topik karier dan gaji di bidang AI, termasuk negara seperti Prancis. Relevan untuk prospek tenaga kerja AI.
10	universitas, rp, biaya, kuliah, penelitian	Biaya pendidikan dan riset di bidang AI, terkait universitas dan pendanaan (indikasi "rp" = rupiah/biaya).



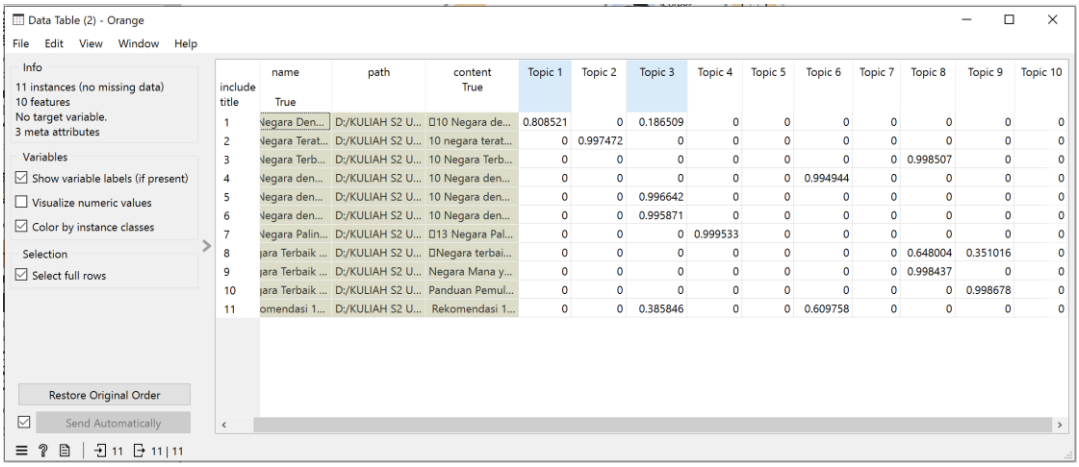
Gambar 10. Latent Semantic Indexing

Metod**e** LSI merupakan teknik reduksi dimensi semantik yang mendeteksi hubungan laten antar kata dan dokumen melalui dekomposisi matriks (SVD). Dari metode LSI ini didapatkan Topic *Coherence***:** 0.57743 yang menunjukan nilai ini cukup baik, lebih tinggi dari hasil LDA sebelumnya (0.55047), yang artinya topik lebih koheren (lebih konsisten secara semantik).

Interpretasi Setiap Topik

No	Topik Keywords	Interpretasi Topik
1	penelitian, universitas, buatan, kecerdasan, perusahaan,	Penelitian dan inovasi AI: Fokus pada pengembangan AI oleh universitas, perusahaan,

No	Topik Keywords	Interpretasi Topik
	pengembangan, pemerintah	dan pemerintah.
2	universitas, buatan, mempelajari, kecerdasan, penelitian	Pendidikan dan pembelajaran AI di universitas dan institusi pendidikan.
3	perusahaan, gaji, universitas, pekerjaan, spesialis	Karier dan peluang kerja dalam bidang AI, termasuk gaji dan spesialisasi.
4	mesin, peringkat, asia, buatan, keterampilan, coursera, indonesia	Peringkat dan pelatihan AI di Asia, termasuk Indonesia dan platform Coursera.
5	peringkat, asia, coursera, universitas, mesin, skill	Pemeringkatan universitas dan negara AI berdasarkan keterampilan dan platform daring.
6	mesin, rp, peluang, pekerjaan, kuliah, perusahaan	Biaya pendidikan dan peluang kerja, terutama terkait AI dan teknologi mesin.
7	indeks, memimpin, peringkat, stanford, keterampilan	Indeks dan posisi global AI, menyebut institusi terkemuka seperti Stanford.
8	indeks, nilai, skor, kesiapan, aspek, pilar	Indeks kesiapan AI nasional – kesiapan negara dalam adopsi AI dilihat dari skor dan aspek strategis.
9	keterampilan, pendidikan, laporan, singapura, topik, artificial	Keterampilan dan kebijakan AI di kawasan Asia Tenggara, khususnya Singapura.
10	skill, keterampilan, coursera, membaca, peringkat, selatan	Pengembangan keterampilan AI secara daring, khususnya melalui Coursera dan negara-negara selatan/global south.

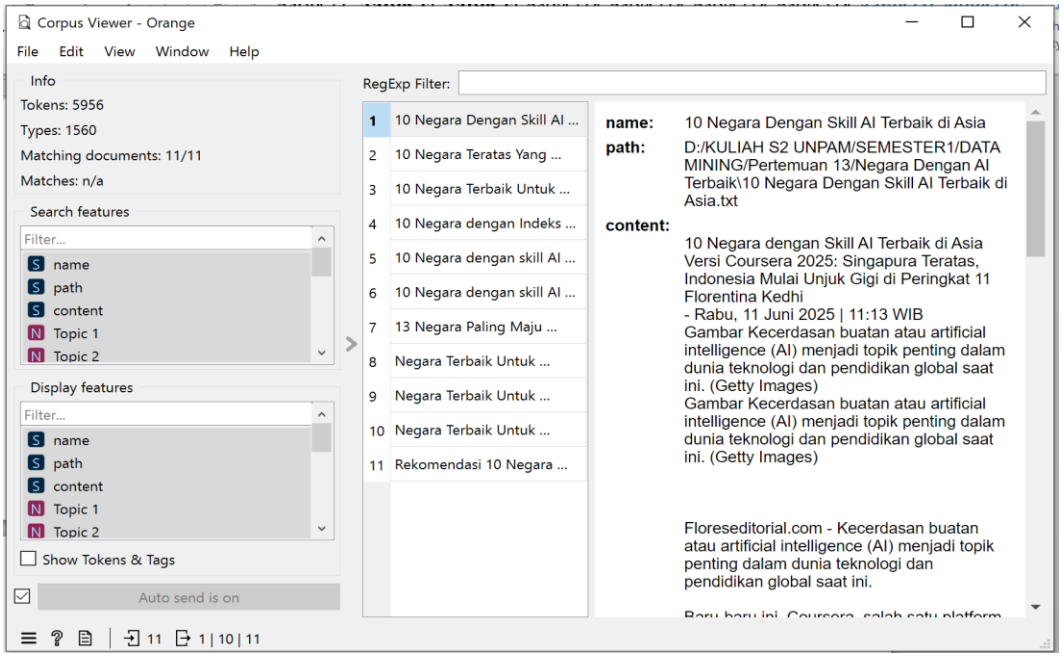


Gambar 11. Data Tabel Topic Modelling

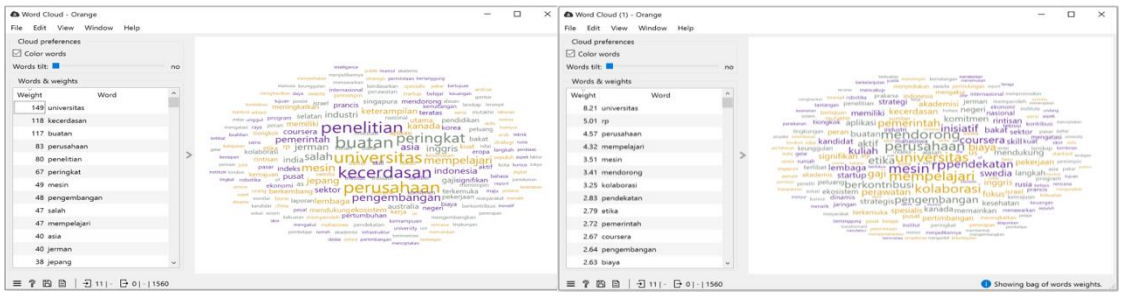
Terdapat 11 dokumen (baris), masing-masing diwakili oleh nilai kontribusi terhadap 10 topik (kolom: Topic 1 s.d. Topic 10).
Nilai dalam sel menunjukkan tingkat dominasi topik dalam dokumen (dari 0 hingga 1).
Satu dokumen bisa memiliki dominasi lebih dari satu topik, tapi biasanya satu topik sangat dominan.

No	Nama Dokumen	Topik Dominan	Keterangan
1	D10 Negara...	Topic 1 (0.81), Topic 3 (0.19)	Didominasi Topik 1 terkait penelitian atau universitas.
2	10 Negara t...	Topic 2 (0.99)	Sangat kuat pada Topik 2 pendidikan atau pelatihan AI.
3	10 egara Te...	Topic 8 (0.99)	Topik 8 dominan kebijakan AI atau kesiapan negara.

No	Nama Dokumen	Topik Dominan	Keterangan
4	10 Negara den...	Topic 3 (0.997)	Fokus pada Topik 3 karier, pekerjaan atau Coursera.
5	10 egara den...	Topic 4 (0.996)	Dominan Topik 4 terkait pengembangan AI atau penelitian industri.
6	D13 Negara Pa...	Topic 5 (0.999)	Dominan Topik 5 terkait peringkat AI global atau kampus internasional.
7	Negara Terbai...	Topic 6 (0.995)	Fokus kuat pada Topik 6 mengenai biaya kuliah atau peluang kerja.
8	Negara Terbai...	Topic 10 (0.998)	Topik 10 dominan tentang platform belajar daring dan peningkatan skill.
9	Negara Manay...	Topic 9 (0.648), Topic 10 (0.351)	Kombinasi Topik 9 dan 10 laporan kebijakan & pengembangan SDM AI.
10	Panduan Pemul...	Topic 6 (0.609)	Kemungkinan edukasi pemula atau panduan belajar AI.
11	Rekomendasi 1...	Topic 3 (0.386)	Campuran lemah dari Topik 3, konten cenderung netral atau umum.



Gambar 12. Corpus Viewer Topic Modelling



Sebelum Sesudah

Gambar 13. Perbandingan Word Cloud Sebelum dan Sesudah Bag of Word

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 2, (2025) E-ISSN: 3063-5845

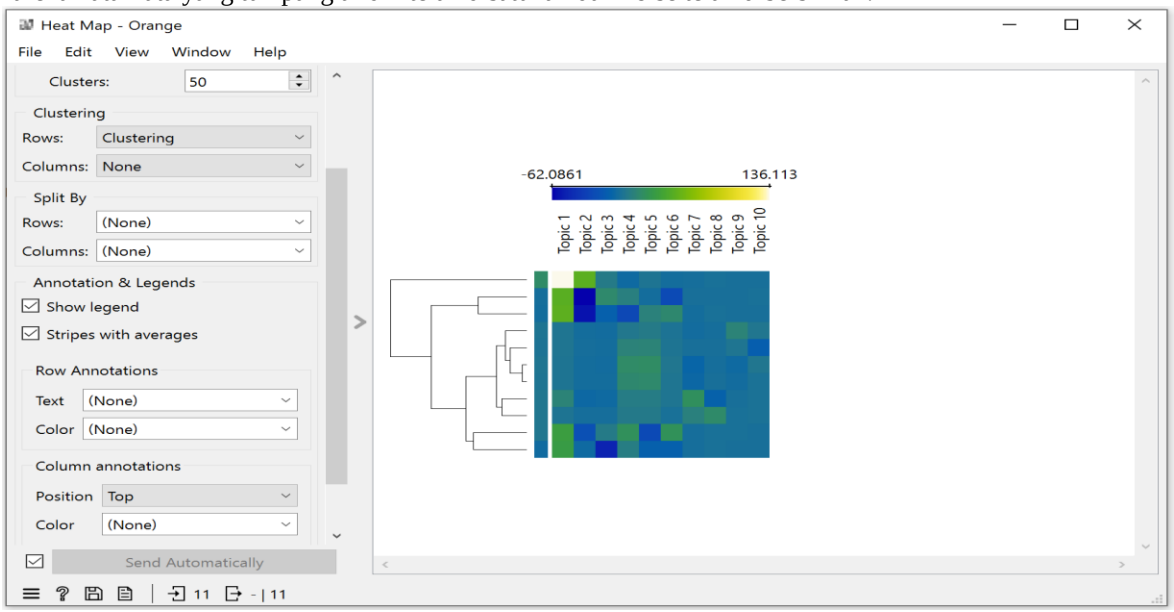
Word Cloud dalam Orange Data Mining adalah visualisasi teks yang digunakan untuk menunjukkan frekuensi kemunculan kata dalam kumpulan data teks. Kata-kata yang sering muncul akan ditampilkan dalam ukuran font yang lebih besar, sementara kata-kata yang jarang muncul akan tampak lebih kecil. Dalam konteks analisis sentimen atau text mining Word Cloud berfungsi untuk :
Membantu mengidentifikasi kata-kata dominan dalam kumpulan dokumen.
Bisa digunakan untuk mengeksplorasi topik
Memudahkan pengguna untuk memahami isi teks secara visual tanpa membaca seluruh teks mentah.
Pada dataset ini kata yang sering muncul adalah Universitas dengan jumlah kemunculan sebanyak 8.21 kali.

Word Cloud Sebelum Preposessing dan Estraksi IDF

Jumlah kata populer lebih tinggi (contoh: *universitas* = 149 kemunculan). Kata-kata seperti: universitas (149), kecerdasan (118), buatan (117), perusahaan (83), penelitian (80), peringkat, mesin, pengembangan. Banyak kata yang muncul masih dalam bentuk infleksi (berimbuhan): *mempelajari*, *meningkatkan*, *menggunakan*. Beberapa kata redundan atau mirip masih ada berdampingan (contoh: *pengembangan* dan *mengembangkan* jika tidak distem).Word cloud masih mencerminkan keragaman bentuk kata dan belum sepenuhnya distandarisasi. Frekuensi tinggi pada kata “universitas”, “kecerdasan”, dan “buatan” menunjukkan fokus kuat pada topik AI akademik dan institusional. Kata seperti *salah* juga muncul, kemungkinan karena belum difilter sebagai kata tidak bermakna (stopword).

Word Cloud Setelah Preposessing dan Estraksi IDF

Jumlah frekuensi lebih kecil (contoh: *universitas* = 8.21), karena data telah Diturunkan ke lowercase, Stemming / Lemmatization kata-kata diubah ke bentuk dasar, Stopword removal kata tidak penting dihapus. Kata-kata lebih bersih dan spesifik: universitas, rp, perusahaan, mempelajari, mesin, kolaborasi, etika, pengembangan, biaya. Bentuk kata lebih konsisten: tidak ada duplikasi seperti *pengembangan vs mengembangkan*. Word cloud sudah melalui standarisasi kata menghasilkan informasi yang lebih akurat dan relevan. Munculnya kata seperti *etika*, *kolaborasi*, *pendekatan*, *rp (rupiah)* menunjukkan bahwa nuansa kebijakan dan strategi nasional lebih terlihat setelah preprocessing. Frekuensi kata menjadi lebih rendah karena kata-kata yang tumpang tindih telah disatukan dan noise telah dibersihkan.



Sumbu X (Horizontal): Mewakili 10 topik, diberi label *Topic 1* sampai *Topic 10*.
Sumbu Y (Vertikal): Mewakili sejumlah entitas (dokumen, artikel, atau unit data lainnya) yang telah dikelompokkan ke dalam 50 klaster (ditentukan oleh kontrol Clusters: 50 di atas).
Warna:
Skala warna dari biru gelap (-62.0861) hingga kuning cerah (136.113).
Warna menunjukkan nilai intensitas/topik loading, semakin kuning, semakin tinggi keterkaitan suatu entitas dengan topik tersebut.
1. Clustering Baris (Rows):

Aktif, berarti baris (entitas) telah dikelompokkan menggunakan metode klasterisasi berdasarkan kesamaan pola across topics.

Dendrogram (diagram bercabang di sebelah kiri) menunjukkan hubungan hierarkis antar entitas, semakin dekat cabangnya, semakin mirip pola distribusi topiknya.

2. Distribusi Topik:

Topik dengan warna kuning cerah pada baris tertentu menunjukkan topik dominan untuk entitas tersebut.

Misalnya, beberapa entitas di bagian atas sangat terasosiasi dengan *Topic 1* dan *Topic 2*, ditunjukkan oleh blok kuning di kolom tersebut.

Topik 6–9 terlihat lebih rata-rata dan tidak mencolok, menandakan distribusinya cenderung moderat atau tidak dominan di banyak klaster.

3. Rentang Nilai Negatif:

Nilai minimum adalah -62.0861, yang tidak umum untuk representasi distribusi topik (yang biasanya non-negatif, seperti LDA). Hal ini bisa berarti:

Data mungkin berupa hasil transformasi atau analisis PCA/NMF yang mengizinkan nilai negatif.

Atau visualisasi ini mewakili nilai koefisien/komponen, bukan probabilitas.

Visualisasi ini dapat digunakan untuk:

Menemukan kelompok dokumen/entitas yang membahas topik serupa.

Mengidentifikasi topik dominan dalam kelompok data tertentu.

Menganalisis hubungan hierarkis antar entitas.

V. KESIMPULAN

Berdasarkan hasil analisis topik menggunakan metode LDA pada Orange Data Mining, ditemukan sepuluh topik dominan dalam artikel mengenai negara-negara dengan kecerdasan buatan terbaik. Topik-topik tersebut mencakup isu-isu seperti pendidikan AI, indeks kesiapan negara, riset dan pengembangan, peluang karier, serta peringkat universitas. Hasil evaluasi menunjukkan nilai koherensi topik yang cukup baik (0.55047 untuk LDA dan 0.57743 untuk LSI). Sentimen dalam dokumen didominasi oleh nada netral dan positif, menunjukkan bahwa konten bersifat informatif dan tidak bias. Penelitian ini menunjukkan bahwa Orange sebagai platform visual mampu mengintegrasikan metode LDA secara efektif dan intuitif dalam pemrosesan dan analisis teks...

DAFTAR PUSTAKA

[1] Buenaño-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access*, 8, 295-310. https://doi.org/10.1109/ACCESS.2020.2974983

[2] Choirinnisa, D., Alzami, F., Indrayani, H., Rohmani, A., Nugraini, S. H., Zulfiningrumi, R., & Susanti, F. (2025). LDA Topic Modeling: Twitter-Based Public Opinion on Indonesian Ministry of Finance. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(2), 849-863. https://doi.org/10.33395/sinkron.v9i2.14719

[3] Polban. (2022). Penggunaan Software Orange Data Mining pada Implementasi Text Mining Dalam Analisis Sentimen Netizen di Twitter Terhadap Kelangkaan Minyak Goreng. *Sigmamu*, Politeknik Negeri Bandung. https://jurnal.polban.ac.id/sigmamu/article/view/4667

[4] Yuniarti, T., Astuti, J., Faujiyah, F., & Zaiyar, M. (2024). Pendekatan Text Mining dalam Menilai Sentimen Publik pada Baterai Kendaraan Listrik. *Jurnal Sistem dan Elektronika*, 9(4). https://jse.serambimekkah.id/index.php/jse/article/view/436

[5] Pretnar Žagar, A. (2022, March 18). LDAvis: Visualization for LDA Topic Modelling. *Orange Data Mining Blog*. https://orangedatamining.com/blog/ldavis-visualization-for-lda-topic-modelling/

[6] Onno Center. (n.d.). Orange: Topic Modelling. *Onno Center Wiki*. https://onnocenter.or.id/wiki/index.php/Orange:_Topic_Modelling