



Metode *Latent Dirichlet Allocation* Untuk Menentukan Topik Pada Review Drama Korea

Alfun Roehatul Jannah^{a,1,*}, Ria Kristi^{b,2}, Muhammad Habibi^{b,3}

^a Informatika, FTTI Unjaya, Yogyakarta, 55293, Indonesia

^b Informatika, FTTI Unjaya, Yogyakarta, 55293, Indonesia

¹ alfunjannah25@gmail.com*; ² riakristibasri@gmail.com; ³ muhammadhabibi17@gmail.com

* corresponding author

ABSTRACT

ARTICLE INFO

Hallyu Wave yang melibatkan penyebaran budaya dan media populer Korea Selatan, telah berkembang pesat selama dua dekade terakhir. Selain industri hiburan seperti K-pop dan K-drama, fenomena ini juga merambah ke sektor makanan dan *K-beauty*. Drama Korea, sebagai inti dari *Hallyu*, telah menjadi fenomena global dengan basis penggemar yang terus berkembang di seluruh dunia. Survei global tahun 2022 menunjukkan bahwa 36 persen responden di 26 negara menganggap drama Korea sangat populer di negara mereka. Di Indonesia, film dan drama Korea tetap menjadi favorit, dengan 72 persen penonton *streaming* memilihnya pada layanan OTT sepanjang 2022. Viu mendominasi sebagai platform streaming drakor paling populer dengan 57 persen penggunaan, diikuti oleh Netflix, Telegram, dan WeTv. Penelitian ini fokus pada analisis data *review* drama Korea dari 2015–2023 menggunakan metode *Latent Dirichlet Allocation* (LDA). Tujuannya adalah memberikan pemahaman mendalam tentang aspek kritis seperti akting, alur cerita, dan sinematografi. Dengan LDA, penelitian ini bertujuan mengidentifikasi topik-topik terkait elemen-elemen tersebut, memberikan wawasan spesifik tentang preferensi penonton. Dari hasil penelitian yang dilakukan, muncul 10 topik ideal dari 20 topik yang ada untuk memastikan konsistensi topik menggunakan *topic coherence*. Dari hasil *topic coherence* untuk 20 topik tersebut, dapat disimpulkan bahwa skor topik keseluruhan untuk topik 10 adalah 0,527 yang memberikan hasil *topic modelling* yang ideal dan sesuai dengan ketentuan.

Article history

Received: 1 April 2024

Revised: 23 May 2024

Accepted: 18 July 2024

Keywords

Drama Korea

Budaya Korea Selatan

k-drama

review

LDA

Latent Dirichlet Allocation

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Pendahuluan

Selama lebih dari dua puluh tahun, pertumbuhan *Hallyu Wave* dapat digambarkan sebagai penyebaran budaya Korea dan media populer ke seluruh dunia [1]. *Hallyu* atau yang dikenal juga dengan *Korea Wave*, menggambarkan fenomena popularitas budaya Korea Selatan. Ini tidak hanya mengacu pada industri hiburan, seperti K-pop dan K-drama, namun juga meluas ke bidang lain seperti makanan dan *K-beauty*.

Drama Korea telah menjadi fenomena global dengan basis penggemar yang terus berkembang di berbagai belahan dunia. Drama korea, juga disebut drakor, adalah sejenis seni yang mengangkat kisah kehidupan manusia dalam bentuk mini seri yang dibuat dalam bahasa Korea. Drama Korea



banyak yang berdasarkan pada cerita romantis, kisah sejarah, thriller, komedi, keluarga, atau campuran dari satu atau lebih genre. Banyak aktor dan aktris Korea yang mampu menarik perhatian dan mengambil hati orang di seluruh dunia. Para artis Korea, dengan bakat dan penampilan yang luar biasa, telah mampu menggerakkan perkembangan industri hiburan Korea, terutama drama dan film.

Menurut survei tahun 2022 yang telah dilakukan di 26 negara, sekitar 36 persen responden menyatakan bahwa drama Korea (K-drama) sangat populer di negara mereka pada tahun tersebut. Menurut laporan Jakpat "2022 Indonesia Mobile Entertainment & Social Media Trends", drama dan film Korea Selatan tetap menjadi pilihan streaming film masyarakat Indonesia dengan persentase sebesar 72% pada layanan OTT (*over the top*). [2]. Pada tahun 2022, survei JakPat menemukan bahwa platform streaming adalah yang paling banyak digunakan oleh orang Indonesia untuk menonton drakor. Viu menjadi platform yang paling banyak digunakan dengan 57%, Netflix berada di peringkat kedua dengan 54%, dan Telegram dan WeTv berada di peringkat ketiga dan keempat [3]. Dari sini penonton bisa secara aktif memberikan ulasan atau *review* langsung di platform ataupun melalui media sosial. Hal ini bisa menjadi rekomendasi personal dari pengguna dan bahan pertimbangan bagi pengguna lain untuk menonton drama.

Dengan banyaknya *review* drama Korea, sulit untuk menemukan topik yang tersembunyi, sehingga diperlukan *topic modelling*. Menemukan topik tersembunyi dapat memberikan manfaat signifikan dalam berbagai aspek seperti memahami preferensi penonton dan sentimen umum terhadap berbagai elemen drama, misalnya plot, karakter, dan kualitas produksi. Dalam penelitian ini, *Latent Dirichlet Allocation* (LDA) digunakan untuk memberikan pemahaman mendalam terhadap aspek-aspek kritis seperti akting, alur cerita, dan sinematografi. Dengan LDA, dapat diidentifikasi dan dibedakan topik-topik yang berkaitan dengan elemen-elemen tersebut, memberikan wawasan yang lebih spesifik tentang preferensi penonton. Dibandingkan metode lain seperti K-Means, NMF, PCA, dan LSA, LDA lebih efektif dalam topik modeling teks karena efisiensinya dan kemampuan untuk menghasilkan distribusi probabilitas topik yang kuat, menjadikannya pilihan utama untuk menganalisis data.

Penelitian serupa menggunakan *topic modelling* LDA untuk mengidentifikasi topik yang sering dibicarakan pelanggan Netflix pada ulasan Google Play. Data ulasan tersebut nantinya dimanfaatkan untuk mengekstraksi tren topik yang ada [4]. Penelitian selanjutnya menggunakan data *review* film China dari Maoyan dan Douban. Pemodelan topik LDA dapat menganalisis topik laten dalam data, memenuhi kebutuhan pelanggan dan meningkatkan pengaruh film China [5]. Selanjutnya, penelitian dilakukan dengan menggunakan pemodelan topik *Latent Dirichlet Allocation* (LDA), yang kemudian diterapkan pada sistem berbasis web. Portal berita Detik.com menyediakan 300 data teks berita, dengan 30 kata per topik [6].

Penelitian ini akan menggunakan data *review* drama Korea dari tahun 2015–2023. Data ini berasal dari platform penyedia data publik open source Kaggle.com. Penelitian ini akan mengelompokkan drama korea berdasarkan elemen penting seperti akting, alur cerita, dan sinematografi melalui pemodelan topik *Latent Dirichlet Allocation* (LDA).

2. Metode Penelitian

Penelitian ini menggunakan teknik studi literatur *text mining* dengan metode *Topic Modelling*, yaitu *Latent Dirichlet Allocation* untuk menemukan topik tersembunyi dalam sebuah *review* film. LDA adalah metode statistik yang menggunakan sebagai model untuk menganalisis dokumen [7]. Sementara menurut Blei, LDA adalah model yang menghasilkan probabilitas dari sekumpulan artikel yang disebut korpus. Pada dasarnya, setiap dokumen terdiri dari penyajian berbagai topik secara acak, dengan setiap topik memiliki karakter tertentu yang didasarkan pada sebaran kata yang ada di dalamnya [8]. Menurut Campbell bahwa LDA menghasilkan daftar topik yang luas untuk setiap dokumen, yang memungkinkan penggunaan LDA untuk merangkum, mengelompokkan, menghubungkan, atau mengolah jumlah data yang sangat besar [9]. LDA menghasilkan daftar subjek

yang diberi bobot pada setiap dokumen, yang membuatnya berguna ketika digunakan untuk meringkas, klusterisasi, menghubungkan, atau memproses data [10].

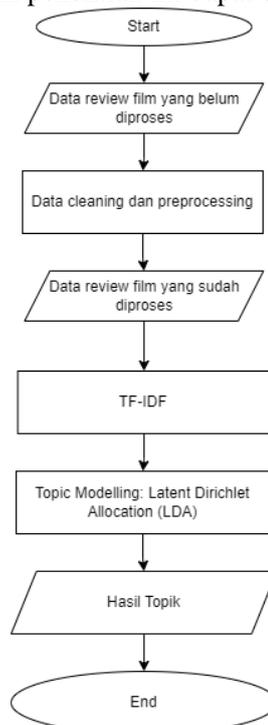
Untuk mencapai hasil yang diinginkan, data yang digunakan dalam penelitian ini berasal dari Kaggle yang bersifat *open source* dan akan diproses melalui proses *preprocessing*. Kaggle adalah *platform online* untuk komunitas *data science* yang menyediakan akses ke berbagai dataset, termasuk *review drama Korea*. Data *review* ini diperoleh melalui *web scraping* atau API dari situs *review* dan *platform streaming*. Setelah dikumpulkan, data tersebut diolah melalui pembersihan, tokenisasi, penghapusan *stop words*, dan *lemmatization*.

Selanjutnya, LDA digunakan untuk menemukan topik tersembunyi dengan membentuk model yang mengidentifikasi kelompok kata yang sering muncul bersama-sama. Hasil analisis ini membantu memahami preferensi penonton, merancang strategi pemasaran yang efektif, dan memberikan masukan untuk meningkatkan kualitas produksi drama. Langkah krusial dalam proses ini adalah penentuan jumlah optimal topik menggunakan perpustakaan seperti Gensim, diikuti dengan evaluasi menggunakan *coherence score*. *Coherence score* digunakan untuk mengukur interpretabilitas dan konsistensi topik berdasarkan hubungan antar kata-kata. Hal ini memastikan bahwa model LDA yang dipilih menghasilkan topik-topik yang bermakna dan relevan dengan data *review drama Korea* yang akan dianalisis.

Setelah mengaplikasikan model LDA untuk menghasilkan 20 topik utama dari *dataset review drama Korea*, langkah berikutnya adalah visualisasi tiap topik dalam bentuk *wordcloud*. *Wordcloud* digunakan untuk merepresentasikan kata-kata yang paling sering muncul dalam setiap topik dengan ukuran kata yang lebih besar menunjukkan frekuensi yang lebih tinggi. Untuk membangun *wordcloud*, kami menggunakan perpustakaan Python seperti *wordcloud* dan *matplotlib*. Setiap kata dalam topik diekstraksi dan dianalisis frekuensinya, kemudian ditampilkan secara visual dalam bentuk *wordcloud* untuk memudahkan pemahaman dan interpretasi terhadap topik-topik yang dihasilkan dari analisis LDA.

Penelitian ini menggunakan laptop yang memiliki spesifikasi yang cukup untuk menjalankan proses pengolahan data dan dapat terhubung ke jaringan internet. Sistem Operasi Windows 10 64-bit, Bahasa Pemrograman Python 3.12.1, Anaconda 2.4.0, dan Jupyter Notebook adalah program-program yang akan digunakan untuk pengolahan data analisis dan *review film*.

Alur penelitian yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Data cleaning dan Preprocessing

Preprocessing yang dilakukan terdiri dari beberapa tahap sebagai berikut.

a. Tokenizing

Tokenizing adalah pemisahan karakter dalam teks, yang dikenal sebagai pemisah kata atau tidak.

b. Stopword Removal

Stopwords juga telah digunakan bersama dengan kata-kata yang panjangnya lebih pendek dari 3 karakter untuk mengurangi beberapa kata yang menyimpang. Tabel 1 berisi daftar kata yang digunakan.

Tabel 1. Daftar Stopword Removal

| Daftar Stopword Removal |
|---|
| ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"] |

2.2. TF-IDF

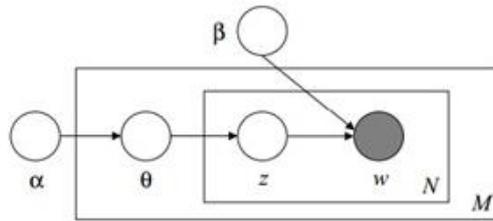
Setelah tahap preprocessing selesai, kumpulan kata dari data review film diberi nilai atau bobot untuk mengetahui seberapa penting kata tersebut. Proses ini menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) untuk menganalisis dan mengevaluasi kata-kata. TF-IDF adalah teknik untuk mengevaluasi pentingnya sebuah kata dalam dokumen relatif terhadap kumpulan dokumen lainnya. TF-IDF menggabungkan frekuensi kemunculan kata dalam dokumen (TF) dan seberapa jarang kata tersebut muncul di seluruh dokumen (IDF). Dalam penelitian analisis review drama Korea, TF-IDF digunakan untuk mengidentifikasi kata-kata yang paling signifikan, mengurangi kebisingan dari kata-kata umum, dan menyusun fitur numerik untuk analisis lebih lanjut. Hal ini membantu meningkatkan akurasi model dan memberikan wawasan mendalam tentang tema dan sentimen yang diungkapkan oleh penonton.

2.3. Topic Modelling

Setelah pembobotan data selesai, pemodelan topik akan dilakukan menggunakan library sklearn algoritma LDA. Kemudian, model LDA akan dilatih dan matriks distribusi topik akan dibuat untuk setiap dokumen.

LDA menghasilkan daftar topik yang diberi bobot pada setiap dokumen, yang membuatnya berguna untuk meringkas, klusterisasi, menghubungkan, atau memproses data. Metode distribusi yang digunakan untuk mendapatkan distribusi topik per dokumen disebut distribusi Dirichlet, dan hasilnya digunakan untuk mengatur kata-kata pada dokumen untuk berbagai topik. Dalam LDA, dokumen adalah objek yang dapat diamati, sedangkan topik, distribusi topik per dokumen, dan penggolongan setiap kata pada topik per dokumen adalah struktur tersembunyi yang tidak dapat diamati secara manual oleh manusia. Oleh karena itu, algoritma ini disebut Latent Dirichlet Allocation [10].

Contoh visualisasi LDA dapat dilihat di Gambar 2.



Gambar 2. Visualisasi LDA

Visualisasi model di atas menunjukkan bahwa model LDA memiliki tiga tingkatan. Parameter α dan β menentukan distribusi topik pada tingkatan corpus, yaitu kumpulan dari dokumen M . Nilai α menunjukkan bahwa jumlah topik yang dibahas dalam dokumen semakin banyak, sedangkan nilai β menunjukkan bahwa jumlah kata yang dibahas dalam topik semakin banyak. Nilai β juga menunjukkan bahwa jumlah kata yang dibahas dalam topik meningkat seiring dengan jumlah kata yang dibahas dalam dokumen. Variabel θ_m , yang berada di tingkat dokumen (M), menunjukkan distribusi topik untuk dokumen tertentu. Nilai θ lebih tinggi menunjukkan bahwa dokumen mengandung lebih banyak topik, sedangkan nilai θ lebih rendah menunjukkan bahwa dokumen semakin spesifik pada topik tersebut. Variabel tingkat kata (N), Z_n , dan W_n adalah variabel Z yang menunjukkan topik kata tertentu dalam dokumen, sedangkan variabel W menunjukkan kata yang berkaitan dengan topik tertentu dalam dokumen [11].

3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan tujuan untuk menggali informasi mengenai topik tertentu dalam data *review* drama Korea. Hasil analisis topik yang dihasilkan dari penerapan metode algoritma LDA diperoleh. Sampel data yang digunakan terdiri dari 10,618 ulasan drama Korea dari tahun 2015 hingga 2023.

3.1. Data cleaning dan preprocessing

Sebelum memulai text mining, data yang sudah ada diproses terlebih dahulu. Proses ini menggunakan *Lemmatizer* dan beberapa kolom juga disingkirkan seperti kolom *'user_id'*, *'story_score'*, *'acting_cast_score'*, *'music_score'*, *'rewatch_value_score'*, *'overall_score'*, *'ep_watched'*, *'n_helpful'*. Hal ini dapat dilihat di Gambar 3.

| | title | review_text |
|-------|---------------------------------|---|
| 0 | Sing My Crush | the Best Song from my Crush This will now ran... |
| 1 | Happy Merry Ending | I'm Happy and Merry That It Ended Quickly The ... |
| 2 | Duty After School: Part 2 | This PART 2 Should Never Have Happened If you... |
| 3 | Our Dating Sim | I want to play this dating sim! I want to pla... |
| 4 | The Director Who Buys Me Dinner | Half-Cooked, Dont Eat A hard pass, unless som... |
| ... | ... | ... |
| 10620 | Heart to Heart | it was a really good drama. highly recommended... |
| 10621 | Kill Me, Heal Me | The only flaws I found in this drama were the ... |
| 10622 | Kill Me, Heal Me | If you're looking for a well written drama abo... |
| 10623 | Kill Me, Heal Me | I started watching this drama, because of the ... |
| 10624 | Kill Me, Heal Me | I really just don't understand the hype for th... |

Gambar 3. Kolom Data Yang Dipilih

Hasil dari *Data cleaning* dapat dilihat pada Gambar 4.

| | title | review_cleaned_text |
|-------|---------------------------------|---|
| 0 | Sing My Crush | Best Song Crush This rank favorite drama ever... |
| 1 | Happy Merry Ending | Happy Merry That Ended Quickly wedding singing... |
| 2 | Duty After School: Part 2 | This PART Should Never Have Happened loved fir... |
| 3 | Our Dating Sim | want play dating want play dating Dating eight... |
| 4 | The Director Who Buys Me Dinner | Half-Cooked Dont hard pas unless someone buy d... |
| ... | ... | ... |
| 10614 | Heart to Heart | really good drama highly recommended liked wri... |
| 10615 | Kill Me, Heal Me | flaw found drama following flashback MANY FLAS... |
| 10616 | Kill Me, Heal Me | looking well written drama mental illness woul... |
| 10617 | Kill Me, Heal Me | started watching drama main couple watched Sec... |
| 10618 | Kill Me, Heal Me | really understand hype could excited watch fou... |

Gambar 4. Hasil Data Cleaning

3.2. TF-IDF

Hasil pembobotan kata menggunakan metode TF ditunjukkan pada Gambar 5 dan 6.

| | |
|---------------|---------------------|
| (10619, 1000) | 0.05644126228916861 |
| (0, 622) | 0.08170356852138365 |
| (0, 951) | 0.17529825352920317 |
| (0, 234) | 0.05276094024135093 |
| (0, 359) | 0.0969915535976581 |
| (0, 322) | 0.0822532834814947 |
| (0, 197) | 0.06539908657527903 |
| (0, 883) | 0.08044137076755475 |
| (0, 95) | 0.0964277226199586 |
| (0, 857) | 0.1037091621995069 |
| (0, 6) | 0.07113545878167042 |
| (0, 303) | 0.09953115646000095 |
| (0, 890) | 0.0955020221997453 |
| (0, 365) | 0.10917243825874634 |
| (0, 498) | 0.10743601213442824 |
| (0, 122) | 0.08313677147504528 |
| (0, 372) | 0.07537852129032768 |
| (0, 539) | 0.09405854129022527 |
| (0, 849) | 0.09196896948819201 |
| (0, 259) | 0.09489025132384965 |
| (0, 954) | 0.10319579288215847 |
| (0, 31) | 0.09037605580774788 |
| (0, 405) | 0.10902892581487238 |
| (0, 429) | 0.24073781894868249 |
| (0, 868) | 0.11444956142068101 |
| (0, 491) | |
| : | : |

Gambar 5. Hasil Perhitungan TF-IDF Awal

| | |
|--------------|----------------------|
| (10618, 911) | 0.0816674486775556 |
| (10618, 656) | 0.09902565764496704 |
| (10618, 513) | 0.16210780104080125 |
| (10618, 205) | 0.06666048796904196 |
| (10618, 637) | 0.056315054990169325 |
| (10618, 347) | 0.09037290511742326 |
| (10618, 891) | 0.11083661133816666 |
| (10618, 561) | 0.21595705158845477 |
| (10618, 425) | 0.16691926520061054 |
| (10618, 701) | 0.07087534228930563 |
| (10618, 919) | 0.15611322967186123 |
| (10618, 880) | 0.05078026413260584 |
| (10618, 951) | 0.04193311866100597 |
| (10618, 359) | 0.05415750640629047 |
| (10618, 197) | 0.08443050307121318 |
| (10618, 522) | 0.07251933875516303 |
| (10618, 715) | 0.08202275168036283 |
| (10618, 528) | 0.04452344303501151 |
| (10618, 497) | 0.05130826234552314 |
| (10618, 481) | 0.11096649996334874 |
| (10618, 326) | 0.10308080270429935 |
| (10618, 548) | 0.07663608654354355 |
| (10618, 305) | 0.06057394303621666 |
| (10618, 511) | 0.0739835081138169 |
| (10618, 222) | 0.12324878663448469 |

Gambar 6. Hasil Perhitungan TF-IDF Akhir

Dengan mencetak kata-kata dengan nilai IDF terkecil dan terbesar, serta mencetak nilai IDF untuk kata-kata tertentu seperti 'acting' dan 'soundtrack', kita dapat mendapatkan wawasan tentang kata-kata yang paling umum dan paling jarang muncul dalam dataset teks tersebut yang dapat dilihat pada gambar 7.

```

nilai IDF terendah (pada indeks 0): drama
nilai IDF tertinggi (pada indeks-1): zombie
acting: 1.9027296168716215
soundtrack: 4.100883562304474
    
```

Gambar 7. Hasil Perhitungan TF-IDF Untuk Kata 'acting' dan 'soundtrack'

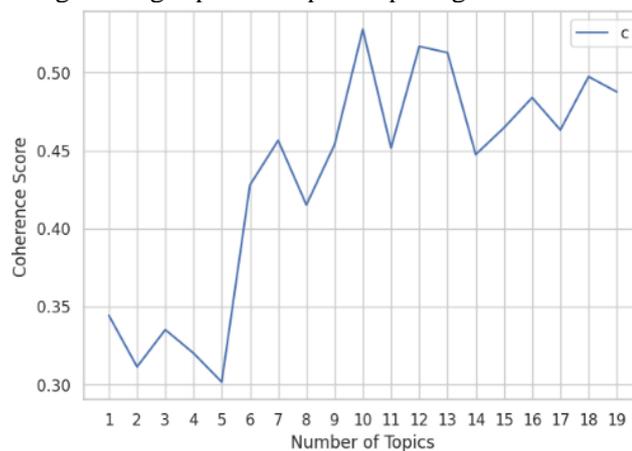
Ini menunjukkan bahwa kata 'drama' adalah kata yang paling umum atau sering muncul dalam dataset teks, sementara kata 'zombie' adalah kata yang lebih jarang muncul dan memiliki kepentingan lebih tinggi.

- Nilai IDF untuk kata 'acting' adalah 1.9027296168716215.
- Nilai IDF untuk kata 'soundtrack' adalah 4.100883562304474.

Nilai-nilai IDF menunjukkan seberapa penting atau uniknya suatu kata dalam dataset teks. Semakin tinggi nilai IDF, semakin jarang kata tersebut muncul di seluruh dataset, dan semakin tinggi bobotnya dalam perhitungan TF-IDF.

3.3. Penjelasan Topic Coherence

Pada tahap evaluasi *topic coherence*, topik terbaik dipilih dari dua puluh topik dari hasil analisis. Skor *coherence* untuk masing-masing topik ditampilkan pada gambar 8.

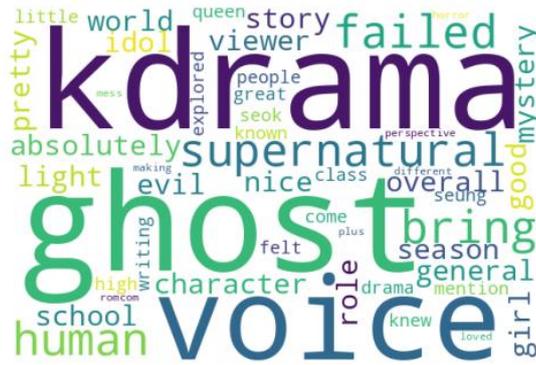


Gambar 8. Grafik Diagram pada *Topic Coherence*

Topik 10 memiliki nilai topik sebesar 0.5277201774702263, menciptakan hasil pemodelan topik yang paling optimal, seperti yang ditunjukkan dalam grafik di atas. Sementara itu, jumlah topik dapat diklasifikasikan menurut nilai koherensi mereka, yang dapat dilihat dalam Tabel 2.

Tabel 2. Nilai *Coherence Score*

| Number of Topics | Coherence Score |
|------------------|---------------------|
| 1 | 0.344278294517525 |
| 2 | 0.3113043321868153 |
| 3 | 0.33512279610165424 |
| 4 | 0.3200611800109456 |
| 5 | 0.30158471077579996 |



Gambar 10. Wordcloud Topik ke-1

Menurut Gambar 10, kata-kata yang sering muncul dalam topik 1 termasuk “*k drama*”, “*ghost*”, “*voice*”, “*supernatural*”, “*bring*”, dan “*human*”. Oleh karena itu, dapat disimpulkan bahwa segmen topik 1 membahas *k drama* yang mengangkat cerita dengan tema dan elemen supranatural, terutama yang berkaitan dengan hantu dan suara.



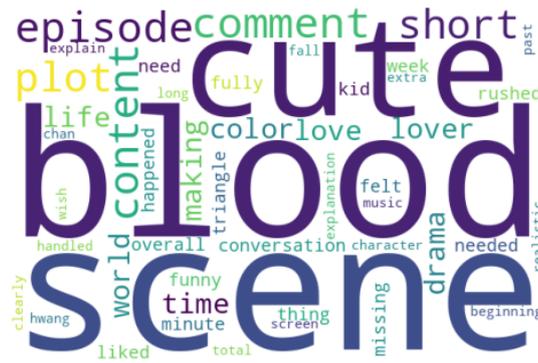
Gambar 11. Wordcloud Topik ke-2

Menurut ilustrasi pada Gambar 11, kata-kata yang sering muncul dalam topik 2 termasuk “*patient*”, “*brain*”, “*case*”, “*good*”, “*sweet*”, dan “*room*”. Oleh karena itu, dapat disimpulkan bahwa segmen 2 membahas aspek positif atau penuh empati dari perawatan pasien atau cerita yang berkaitan dengan situasi medis.



Gambar 12. Wordcloud Topik ke-3

Menurut Gambar 12, kata-kata yang sering muncul dalam topik 3 adalah “*drama*”, “*cancel*”, “*helpful*”, “*short*”, “*really*”, dan “*good*”. Oleh karena itu, dapat disimpulkan bahwa topik ini berkaitan dengan drama yang berfokus pada pembatalan acara, evaluasi tingkat kemanfaatan, durasi yang singkat, dan evaluasi kualitas drama.



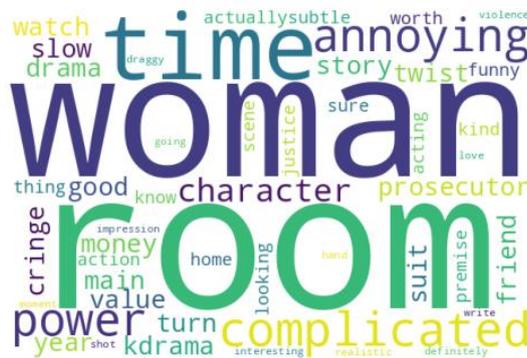
Gambar 13. Wordcloud Topik ke-4

Topik 4 mengandung kata-kata yang sering muncul seperti “blood”, “scene”, “cute”, “episode”, “comment”, dan “content”, seperti yang ditunjukkan pada Gambar 13. Oleh karena itu, dapat disimpulkan bahwa topik 4 berkaitan dengan respons terhadap konten drama yang berkaitan dengan episode dengan adegan dramatis dan momen yang menggemaskan.



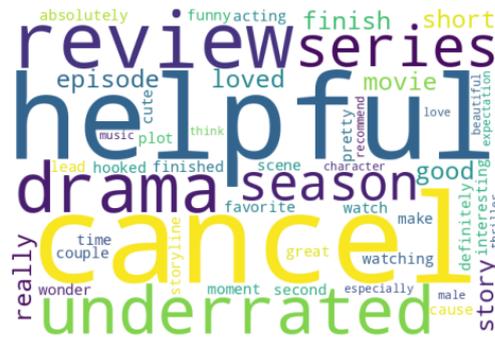
Gambar 14. Wordcloud Topik ke-5

Menurut Gambar 14, kata-kata yang sering muncul dalam topik 5 termasuk “season”, “humor”, “bullying”, “husband”, “doctor”, dan “medical”. Oleh karena itu, dapat disimpulkan bahwa topik ini membahas drama yang berkaitan dengan musim tertentu dengan elemen tema humor, masalah pelecehan, dan aspek-aspek medis atau hubungan antar karakter.



Gambar 15. Wordcloud Topik ke-6

Dalam topik 6, kata-kata yang sering muncul termasuk “room”, “woman”, “time”, “complicated”, “annoying”, dan “power”, seperti yang ditunjukkan pada Gambar 15. Oleh karena itu, dapat disimpulkan bahwa topik 6 membahas drama yang melibatkan karakter perempuan dalam suatu ruangan atau lingkungan dengan dinamika waktu yang kompleks dan melibatkan elemen kekuatan.



Gambar 16. Wordcloud Topik ke-7

Menurut Gambar 16, kata-kata yang sering muncul dalam topik 7 termasuk “cancel”, “helpful”, “review”, “drama”, “underrated”, dan “series”. Oleh karena itu, kita dapat mengambil kesimpulan bahwa topik 7 membahas tentang pembatalan acara, penilaian tingkat kemanfaatan, dan penilaian dan penghargaan untuk drama yang mungkin tidak terkenal tetapi menerima ulasan positif.



Gambar 17. Wordcloud Topik ke-8

Menurut Gambar 17, kata-kata yang sering muncul dalam topik 8 termasuk “mini”, “drama”, “cute”, “idol”, “short”, dan “romance”. Dengan demikian, topik ini membahas pembuatan serial drama mini yang menggabungkan elemen keceriaan atau menggemaskan, romantis, dan keterlibatan tokoh idol atau elemen yang terkait dengan budaya K-pop.



Gambar 18. Wordcloud Topik ke-9

Dalam topik 9, kata-kata yang sering muncul termasuk “dramas”, “hyun”, “ridiculous”, “like”, “king”, dan “historical”. Oleh karena itu, dapat disimpulkan bahwa topik ini membahas drama sejarah dengan sentuhan kerajaan serta diskusi tentang aspek-aspek yang dianggap tidak masuk akal atau lucu dalam drama tersebut.

3.5. Hasil Analisis per topik

Untuk membuat pembahasan atau kesimpulan dari kumpulan kata-kata dalam setiap topik yang dihasilkan dari analisis menggunakan LDA pada review drama Korea, langkah pertama adalah mengidentifikasi kata-kata kunci yang mewakili tema atau aspek utama dari masing-masing topik. Misalnya, dalam topik yang menyoroti karakter dan cerita drama, kata-kata seperti “*character*”, “*story*”, dan “*episode*” dominan. Kemudian, interpretasikan kata-kata tersebut dalam konteks yang lebih luas untuk memahami diskusi tentang pengembangan karakter atau plot cerita dalam drama. Analisis hubungan antar kata juga penting untuk memperjelas kontribusi masing-masing kata terhadap pembentukan topik. Metodologi yang digunakan, seperti LDA, telah terbukti efektif dalam mengungkap struktur topik dari teks, sementara pendekatan lain seperti analisis sentimen dan teknik *word embeddings* (seperti Word2Vec) dapat memberikan wawasan tambahan tentang opini atau sentimen yang terkandung dalam ulasan. Dengan demikian, penggabungan langkah-langkah ini dengan literatur yang mendukung memungkinkan pembahasan yang komprehensif dan informatif tentang setiap topik yang diteliti.

Berdasarkan temuan penelitian ini, metode topic modelling digunakan untuk menganalisis topik *review* drama korea dari tahun 2015–2023. Topik-topik tersebut ditampilkan dalam tabel 3 dengan 10 jumlah topik.

Tabel 3. Kata-kata per topik

| No Topik | Kata-kata | Pembahasan |
|----------|--|---|
| 0 | <i>drama, character, story, really, like, episode, love, watch, good, time</i> | mengapresiasi karakter, aspek-aspek pengembangan cerita, dan episode tertentu |
| 1 | <i>kdrama, ghost, voice, supernatural, bring, failed, human, absolutely, season, story</i> | kdrama yang mengangkat cerita dengan tema dan unsur supranatural, terutama berkaitan dengan tema hantu dan suara |
| 2 | <i>patient, brain, case, good, sweet, room, like, body, told, episode</i> | aspek positif atau penuh empati dalam penanganan pasien atau dalam cerita yang melibatkan situasi medis |
| 3 | <i>drama, cancel, helpful, short, really, good, review, watch, story, cute</i> | pembahasan terkait pembatalan acara, penilaian terhadap tingkat kemanfaatan, durasi yang singkat, dan penilaian kualitas dari drama tersebut. |
| 4 | <i>blood, scene, cute, episode, comment, content, short, plot, color, time</i> | tanggapan terhadap konten drama tentang episode dengan adegan yang dramatis dan momen yang menggemaskan. |
| 5 | <i>season, humor, bullying, husband, doctor, medical, loved, wife, love, drama</i> | drama yang melibatkan <i>season</i> tertentu dengan tema atau unsur humor, isu <i>bullying</i> , dan aspek-aspek medis atau hubungan antar karakter. |
| 6 | <i>room, woman, time, complicated, annoying, power, character, prosecutor, suit, slow</i> | drama yang melibatkan karakter perempuan dalam suatu ruangan atau lingkungan dengan dinamika waktu yang kompleks dan melibatkan elemen kekuatan |
| 7 | <i>cancel, helpful, review, drama, underrated, series, season, episode, loved, good</i> | pembahasan tentang pembatalan acara, evaluasi tingkat kemanfaatan, serta penilaian dan penghargaan untuk drama-drama yang mungkin kurang terkenal, namun dinilai positif oleh sebagian penonton |

respon emosional penonton seperti kegemaran terhadap momen dramatis atau momen menggemaskan dalam sebuah episode. Hal ini dapat dilihat dari kata-kata seperti “*drama*”, “*story*”, “*episode*”, “*good*”, “*like*”, “*watch*”, “*time*”, “*season*”, “*short*”, dan “*cute*” adalah yang paling dominan muncul dari seluruh *wordcloud*.

References

- [1] G. Ganghariya and R. Kanozia, “Proliferation of Hallyu wave and Korean popular culture across the world: A systematic literature review from 2000-2019,” *Journal of Content, Community and Communication*, vol. 10, no. 6. Amity University, pp. 177–207, Jun. 01, 2020. doi: 10.31620/JCCC.06.20/14.
- [2] R. Hasya, “Drama Korea Masih Jadi Favorit Masyarakat Indonesia dalam Streaming Film dan Serial di Tahun 2022,” GoodStats.
- [3] R. Pahlevi, “Bukan Netflix, Penonton Drakor Indonesia Paling Banyak Nonton Lewat Platform Ini,” databoks.
- [4] G. Rosalinda, R. Santoso, and P. Kartikasari, “PEMODELAN TOPIK ULASAN APLIKASI NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN LATENT DIRICHLET ALLOCATION,” *Jurnal Gaussian*, vol. 11, no. 4, pp. 554–561, Feb. 2023, doi: 10.14710/j.gauss.11.4.554-561.
- [5] L. Zhao, Q. Zhao, and Y. Wang, “Research on Chinese Movie Reviews Based on Latent Dirichlet Allocation Topic Model,” in *Proceedings - 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 46–49. doi: 10.1109/MLBDBI51377.2020.00016.
- [6] S. Karmila, V. Intan Ardianti Prodi Telematika Energi, and V. Intan Ardianti, “METODE LATENT DIRICHLET ALLOCATION UNTUK MENENTUKAN TOPIK TEKS SUATU BERITA,” *Jurnal Informatika & Komputasi*, vol. 16, 2022.
- [7] M. D. R Wahyudi, A. Fatwanto, U. Kiftiyani, and M. Galih Wonoseto, “Topic Modeling of Online Media News Titles during COVID-19 Emergency Response in Indonesia Using the Latent Dirichlet Allocation (LDA) Algorithm,” *Telematika*, vol. 14, no. 2, pp. 101–111, Aug. 2021, doi: 10.35671/telematika.v14i2.1225.
- [8] D. M. Blei, A. Y. Ng, and J. B. Edu, “Latent Dirichlet Allocation Michael I. Jordan,” 2003.
- [9] J. C. Campbell, A. Hindle, and E. Stroulia, “Latent Dirichlet Allocation,” in *The Art and Science of Analyzing Software Data*, Elsevier, 2015, pp. 139–159. doi: 10.1016/B978-0-12-411519-4.00006-9.
- [10] T. K. Kurniasari, W. Maharani, and J. H. Husen, “Analisis Media Sosial Twitter untuk Mengetahui Pengguna Berpengaruh pada Portal Berita di Indonesia dengan Metode TSIM (Topic-based Social Influence Measurement).”
- [11] K. B. Putra and R. P. Kusumawardani, “Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA),” *Jurnal Teknik ITS*, vol. 6, no. 2, Sep. 2017, doi: 10.12962/j23373539.v6i2.23205.