

ANALISIS TREN KONTEN PADA VTUBER INDONESIA MENGGUNAKAN LATENT DIRICHLET ALLOCATION

Akdeas Oktanae Widodo¹, Farhan Septiadi², Nur Aini Rakhmawati³

¹²³Program Studi S1 Sistem Informasi, Institut Teknologi Sepuluh Nopember

Jln. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur 60111

¹ akdeas.19052@mhs.its.ac.id, ² septiadifarhan.19052@mhs.its.ac.id, ³ nur.aini@is.its.ac.id

Abstract

YouTube is evidence of latest digital technology developments in the field of media and entertainment. Not all of youtubers or people who create video content on YouTube express themselves directly, some use intermediaries of two-dimensional or three-dimensional virtual characters created with help of computer software to interact with the audience, which in future is commonly referred to as Virtual Youtuber (VTuber). This vtuber trend began to become famous in 2016 in Japan and is increasing throughout the world, Then a lot of people are interested in becoming part of vtuber as well. This research helps to find out the content topics aired by Famous vtuber using the Latent Dirichlet Allocation (LDA) method. The analysis is done after running the text mining process on 4312 videos from the top 10 vtuber channels in Indonesia. Determination of optimal topic applied LDA based the results of the perplexity value and topic coherence. From the implementation of the LDA method, the results were obtained in the form of five topics that are often broadcast by vtubers, including Minecraft games and reading donations, the Apex Legend game accompanied by collaborations with other vtubers, live video game broadcasts, covers of other singers' songs and finally other multiplayer game shows like Raft or Phasmophobia.

Keywords : *latent dirichlet allocation, topic modeling, virtual youtuber, vtuber*

Abstrak

YouTube merupakan bukti dari perkembangan teknologi digital terbaru dalam bidang media dan hiburan. Tidak semua *youtuber* atau orang yang membuat konten video di YouTube melakukan ekspresi diri secara langsung, ada yang menggunakan perantara karakter virtual dua dimensi atau tiga dimensi yang dibuat dengan bantuan perangkat lunak computer untuk berinteraksi dengan penonton yang kedepannya biasa disebut sebagai *Virtual Youtuber (VTuber)*. Tren *vtuber* ini mulai terkenal pada tahun 2016 di Jepang makin tahun makin meningkat diseluruh dunia, Kemudian Semakin banyak orang-orang menyukai konten yang dibawakan oleh *Vtuber* dan tidak sedikit yang tertarik untuk menjadi bagian dari *vtuber* juga. Untuk membantu hal tersebut penelitian ini membantu untuk mengetahui topik konten yang ditayangkan oleh *Vtuber* Terkenal menggunakan metode *Latent Dirichlet Allocation (LDA)*. Analisis dilakukan setelah menjalankan proses *text mining* terhadap 4312 video dari 10 *channel vtuber* Indonesia teratas. Penentuan topik yang optimal dari LDA yang diterapkan dapat melihat hasil nilai *perplexity* dan *topic coherence*. Dari implementasi metode LDA didapatkan hasil berupa lima topik yang sering ditayangkan oleh *vtuber* antara lain gim Minecraft dan reading donation, gim Apex Legend disertai *collaboration* dengan *vtuber* lain, tayangan siaran langsung *video game*, *cover* dari lagu penyanyi lain dan terakhir tayangan gim *multiplayer* lain seperti Raft atau Phasmophobia.

Kata kunci : *latent dirichlet allocation, topic modeling, virtual youtuber, vtuber*

1. PENDAHULUAN

YouTube merupakan bukti dari perkembangan teknologi digital terbaru dalam bidang media dan hiburan. Platform video tersebut sangat dipopuleri oleh berbagai kalangan

sehingga mendukung orang-orang untuk kontribusi di aplikasi YouTube tersebut dengan mengekspresikan diri seperti membuat serta menayangkan video tentang hal yang disukai, kehidupan sehari-hari, trend terkini dan lain

sebagainya yang mana nanti disebut sebagai *youtuber* [1].

Tidak semua *youtuber* melakukan ekspresi diri secara langsung, ada yang menggunakan perantara karakter virtual 2 dimensi atau 3 dimensi yang dibuat dengan bantuan perangkat lunak komputer untuk berinteraksi dengan penonton di platform YouTube [2]. Penggunaan karakter virtual tersebut mengacu pada karakter di akun tersebut untuk memberikan kesan lucu. Penggambaran karakter yang bersifat lucu tersebut oleh masyarakat Jepang disebut sebagai *Antropomorfisme Moe* [3]. Untuk memberikan kesan lucu kebanyakan karakter virtual ini digambarkan dengan karakter berkelamin perempuan. Istilah untuk *Youtuber* menggunakan karakter virtual ini kelak disebut sebagai *Virtual YouTuber* disingkat *vtuber*.

Vtuber pertama kali muncul pada tahun 2016 di Jepang dan mengadopsi karakteristik visual seperti anime yaitu Kizuna Ai. Disaat yang bersamaan dialah yang mengungkapkan identitas bahwa dia adalah seorang *vtuber* di kanal YouTube-nya sehingga bisa disebut sebagai *vtuber* Pertama. Kemudian selanjutnya kegiatan Kizuna Ai dimanajemen oleh perusahaan agensi *vtuber* yaitu Kizuna AI Inc [4]. Tren *vtuber* ini makin tahun makin meningkat, semakin banyak orang-orang menyukai konten yang dibawa oleh *vtuber* dan tidak sedikit yang tertarik untuk menjadi bagian dari *vtuber* juga. Maka dari itu dengan semakin banyaknya peminat *vtuber* ini, memahami pola dari apa yang ditonton serta disukai penonton merupakan hal penting yang dapat membantu *vtuber* dalam membuat konten yang menarik dan berkualitas.[1].

Uraian di atas menginspirasi peneliti memberikan solusi untuk orang lain yang berminat menjadi *vtuber* atau *vtuber*. Peneliti menggunakan metode analisis *topic modeling* terhadap konten yang dibuat oleh *vtuber* yang sudah terkenal. Pengambilan data diambil berdasarkan konten yang dibuat oleh *vtuber* dengan *subscriber* di atas seratus ribu. *Topic Modelling* ini merupakan pendekatan pada *Text Mining* untuk menemukan data-data pada sebuah informasi berupa text.[1]

Setelah data dikumpulkan penelitian ini akan dilanjutkan ke tahap *Latent Dirichlet Allocation* (LDA) yang digunakan untuk membantu menentukan topik yang sudah ada pada akun *vtuber* di YouTube. LDA ini dilakukan dengan meringkas, melakukan klasterisasi, menghubungkan hingga memproses data agar menghasilkan daftar topik yang diberi bobot untuk setiap informasi. Banyak penelitian yang menggunakan metode LDA dikarenakan saat ini

merupakan metode yang populer, sehingga hal itu menjadi salah satu alasan untuk peneliti menggunakan teknik LDA ini.[1]

Penelitian ini mengikuti referensi dari Alfrida Rahmawati pada tahun 2021 [1], dengan perbedaan objek penelitian. Penelitian ini meneliti tentang tren konten yang diminati oleh *vtuber* Indonesia. Sejauh ini untuk sumber data utama masih sama dengan penelitian sebelumnya, yaitu menggunakan judul video, akan tetapi peneliti juga menyertakan video *archive livestream* sebagai sumber data. Kemudian di analisis dengan metode LDA sama seperti penelitian sebelumnya. Harapan peneliti pada penelitian ini dapat membantu *vtuber* untuk mendapatkan referensi mengenai tren konten apa yang diminati penonton dan ditayangkan oleh *vtuber* lainnya supaya bisa mengoptimalkan penyampaian konten yang akan dibuat.

1.1. YouTube

YouTube merupakan situs web layanan yang dapat memungkinkan pengguna untuk menonton maupun mengunggah video secara daring selama terkoneksi dengan jaringan internet. Berbagai macam konten video yang dapat ditemukan pada platform ini seperti klip film, klip televisi, video musik, vlog, film pendek, video edukasi, video tutorial, hingga konten *gaming*[5].

1.2. Virtual YouTuber

Virtual YouTuber atau yang biasa disebut sebagai *vtuber* merupakan *streamer* dan *vlogger* yang menggunakan karakter 2D maupun 3D yang dihasilkan dari komputer[2]. Sesuai dengan namanya, *vtuber* biasa melakukan *livestream* pada platform YouTube. *Vtuber* yang populer dan banyak peminatnya biasanya dibawa oleh suatu agensi. Fenomena yang biasa terjadi adalah ketika agensi *vtuber* luar negeri yang sudah terkenal namanya membuka cabang di Indonesia, maka *vtuber* yang dibawa nama agensi tersebut biasanya akan bisa cepat meraih popularitas dalam kurun waktu yang singkat[6].

1.3. Topic Modeling

Topic modeling adalah salah satu teknik analisis data yang digunakan untuk mengelompokkan dokumen atau teks menjadi beberapa topik yang berbeda. Tujuan dari *topic modeling* adalah untuk mengidentifikasi kategori atau tema yang muncul dalam sekumpulan dokumen dan menentukan distribusi topik tersebut di antara dokumen. Dengan demikian,

topic modeling merupakan teknik yang berguna untuk mengelompokkan dokumen atau teks menjadi beberapa topik yang terkait dan mengekstraksi informasi dari dokumen yang tidak terstruktur. Teknik ini bermanfaat untuk memudahkan proses pencarian informasi, mengidentifikasi tren atau perkembangan dalam sekumpulan dokumen, dan mengekstraksi fitur yang relevan untuk proses klasifikasi atau *clustering* [7].

1.4. Metode Latent Dirichlet Allocation (LDA)

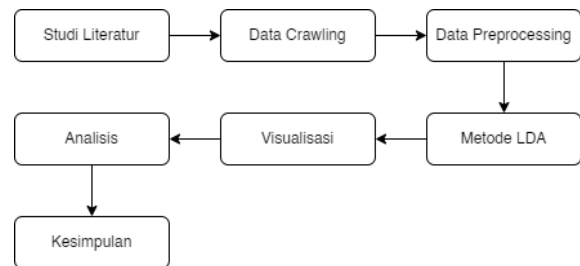
Latent Dirichlet Allocation (LDA) merupakan salah satu metode *topic modeling* yang paling populer saat ini. Seperti yang sudah dijelaskan pada pendahuluan yang mana salah satu pendekatan dengan menggunakan *text mining*. Secara pengertian *text mining* dapat diartikan sebagai penemuan informasi baru yang sebelumnya belum diketahui oleh *computed-machine* dengan cara diotomatiskan untuk ekstraksi informasi dari berbagai sumber dengan kunci dalam prosesnya berupa penggabungan informasi dari berbagai sumber tersebut [8].

Dikarenakan LDA merupakan *unsupervised learning* untuk *topic modeling*, LDA tidak memerlukan *training set*, *tag*, atau *metadata* apapun untuk *learning*, sehingga dokumen tekstual dalam jumlah yang besar dapat dianalisis dalam waktu yang relative singkat [9]. Oleh karena itu, metode LDA ini digunakan untuk analisis *topic modeling* dalam penelitian ini.

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Penelitian ini akan dilakukan melalui dua tahapan untuk dapat mencapai hasil yang sesuai dengan topik dan tujuan dari penelitian ini. Terdapat tiga tahapan pada akuisisi data antara lain studi literatur, pengumpulan data, dan pengolahan data. Kemudian dilanjutkan ke analisis yang mana terdapat dua tahapan yaitu analisis data dan penarikan kesimpulan. Untuk lebih jelasnya, diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.2. Studi Literatur

Tahapan awal dalam penelitian ini yaitu melakukan studi literatur untuk mendapatkan informasi terkait dengan metode maupun alur kerja dari jurnal terdahulu yang masih berkaitan dengan penelitian ini sebagai referensi. Rujukan studi literturnya antara lain artikel ilmiah, jurnal, tugas akhir, dan situs web resmi yang kredibel sehingga dapat mendukung penelitian ini.

2.3. Crawling Data

Selanjutnya akan dilakukan *crawling* data yang mana nanti peneliti akan mengumpulkan data dari website YouTube dengan acuan *hyperlink* untuk pengindeksan pencarian data dalam dokumen per setiap *link* yang ada [1]. *Crawling* merupakan teknik yang digunakan untuk mengumpulkan data yang terdapat dalam sebuah *website*, bekerja secara otomatis yang mana informasi yang didapatkan sesuai dengan kata kunci yang sudah dimasukkan [10]. Data yang dikumpulkan adalah data terkait dengan *vtuber*. Data tersebut berupa topik yang difokuskan kepada *channel* YouTube dengan konten *vtuber*. Peneliti menghimpun sepuluh *channel* YouTube *vtuber* Indonesia dengan *subscriber* tertinggi untuk dijadikan target *crawling*. Selanjutnya akan dilakukan *data mining* untuk mendapatkan data berupa judul video atau siaran langsung, jumlah *subscriber* dan jumlah video atau siaran langsung yang perlu dihimpun untuk proses selanjutnya. Penelitian ini membutuhkan *API key* dan *channel id* dari masing-masing akun *channel* YouTube *vtuber* Indonesia tersebut untuk dijadikan sebagai bahan *scrapping* dari *crawling website* yang akan dilakukan.

2.4. Data Preprocessing

Data yang sudah didapatkan akan diproses sebelum digunakan, tahap ini disebut *Data Preprocessing* yang mana merupakan Langkah untuk menghasilkan data yang siap digunakan

dalam pelabelan serta permodelan data sehingga tahap ini bertujuan untuk pembersihan kata yang tidak perlu atau tidak maknawi [11]. Proses *cleaning* menjadikan data lebih teragregasi sehingga dapat digunakan dengan mudah waktu ditahap selanjutnya. Setiap data yang memiliki jenis yang sama akan dimasukkan kedalam suatu indeks yang bertujuan membedakan antar dokumen. Dari sekumpulan data yang sudah didapatkan tersebut akan dilakukan proses tokenisasi yang bertujuan untuk memisahkan teks dalam unit yang lebih kecil, biasanya disebut sebagai token [12].

Pada tahap ini juga dilakukan *remove stopwords* yang mana nanti jika dalam suatu kalimat dari hasil yang sudah didapatkan akan dianggap tidak penting [13], lalu dilewati waktu penerapan LDA nya.

2.5. Metode LDA

Sebelumnya data sudah teragregasi kedalam beberapa indeks, setelah itu baru kita bisa menerapkan LDA kedalam set datanya. Penerapan LDA bertujuan untuk memperoleh persebaran data yang membentuk proyeksi linear yang akan menghasilkan topik atas suatu dokumen [1]. Penerapan metode ini akan dibagi menjadi beberapa tahap diawali dengan mengisi parameter yang akan digunakan sebagai batasan dari data yang digunakan, lalu menerapkan *semi random distributor* yang mana didasari atas metode distribusi *Dirichlet*, lalu dilakukannya iterasi untuk memperlihatkan parameter yang nantinya dapat menentukan distribusi topik berupa kata dari dokumen yang diproses [1].

Penentuan banyaknya topik optimalnya dapat menggunakan perhitungan *perplexity* yang mana itu dapat menentukan kemungkinan dari logs teks dokumen yang tidak kasat mata. Hasil yang didapatkan dari nilai *perplexity* ternyata rendah, maka model yang dibentuk tergolong baik [1]. Ada kalanya cara *perplexity* ini tidak berkolerasi dengan interpretasi manusia dengan baik.

2.6. Visualisasi

Setelah data yang didapatkan dari penerapan metode LDA ini menjadi data mentah, kemudian akan dilanjutkan dengan visualisasi dengan grafik berupa *bar-chart* untuk mempermudah proses analisis nantinya [1]. Visualisasi data ini akan dilakukan dapat mengetahui seberapa besar persebaran antar data.

2.7. Analisis

Analisis data akan dilakukan dengan metode *topic modeling*, yang mana dapat mengelompokkan topik tertentu berdasarkan kemiripan pola atau kesamaan kata kunci yang sudah ditemukan. *Topic modeling* sendiri adalah metode statistik untuk menemukan pola dan tema dalam sebuah corpus dokumen [14], sehingga relevan bagi penelitian ini untuk menemukan topik paling sering ditayangkan dari *channel vtuber* tersebut.

Hasil visualisasi data sebelumnya akan diikutkan dalam analisis terkait proporsi mengenai topik yang ditemukan. Hasil analisis ini nanti akan dievaluasi untuk menemukan topik yang dominan dalam suatu konten dan konten tersebut yang dominan dibahas disetiap *channel*.

2.8. Kesimpulan

Data telah melewati tahap analisis sehingga kita dapat membuatnya menjadi kesimpulan dari konten sepuluh *channel* yang dijadikan subjek dalam penelitian ini. Didapatkan nanti lima topik teratas yang menjadi pokok bahasan utama atau yang sering diunggah atau disiarkan pada *channel vtuber* tersebut. Kesimpulan ini nantinya juga mencerminkan topik konten yang dinikmati oleh para penonton *vtuber*.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Crawling Data

Setelah studi literatur selesai dilakukan, selanjutnya adalah *crawling data*. Data yang dipakai pada penelitian ini merupakan data *channel* YouTube *vtuber* Indonesia yang memuat jumlah *subscriber*, nama *channel*, hingga judul dan deskripsi video. Data yang diambil merupakan data yang terakhir diperbarui per tanggal 27 September 2022.

Data diambil berdasarkan *channel* YouTube *vtuber* Indonesia dengan jumlah *subscriber* terbanyak. Berdasarkan kriteria tersebut, diambil sepuluh sampel data untuk diteliti. Daftar *channel* yang diambil dapat dilihat pada Tabel 1.

TABEL 1. DAFTAR KANAL VTUBER INDONESIA DENGAN JUMLAH *SUBSCRIBER* TERBANYAK

No	Nama <i>Channel</i>	Jumlah <i>Subscriber</i> (k=ribu)
1	Kobo Kanaeru	1430k
2	Moona Hoshinova	1190k
3	Kureiji Ollie	1180k

4	Ayunda Risu	732k
5	Pavolia Reine	693k
6	Mythia Batford	660k
7	Andi Adinata	640k
8	Airani Iofifteen	628k
9	Vestia Zeta	563k
10	Any Melfissa	562k

Dapat dilihat pada Tabel 1 bahwa *subscriber vtuber* Indonesia paling banyak saat ini adalah Kobo Kanaeru dengan total 1,43 juta *subscriber*. Sedangkan untuk yang paling sedikit di antara sepuluh terbanyak adalah Anya Melfisa dengan jumlah 562 ribu *subscriber*.

TABEL II. DAFTAR JUMLAH VIDEO VTUBER INDONESIA

No	Nama Channel	Jumlah Video
1	Kobo Kanaeru	149
2	Moona Hoshinova	478
3	Kureiji Ollie	582
4	Ayunda Risu	576
5	Pavolia Reine	664
6	Mythia Batford	304
7	Andi Adinata	503
8	Airani Iofifteen	456
9	Vestia Zeta	189
10	Any Melfissa	412
	Total	4313

Setelah didapatkannya API key dan *channel id*, maka dapat diambil judul dan deskripsi yang mengandung waktu pengunggahan video serta keterangan dari setiap *channel vtuber* tersebut. Setelah beberapa data antara lain judul, durasi, deskripsi, dan waktu pengunggahan didapatkan kemudian disatukan dalam bentuk table yang disimpan dalam bentuk .csv sebagai bahan input dari proses analisis berikutnya dan diunggah melalui akun Zenodo [15].

Dapat dilihat pada Tabel 2 yang merupakan jumlah video dari beberapa *channel YouTube vtuber*. Didapatkan hasil yang beragam pada masing-masing *channel*. Terdapat jumlah video yang paling banyak yaitu 664 video pada *channel Pavolia Reine* dan yang paling sedikit 149 video pada *channel Kobo Kanaeru*. Total keseluruhan dari sepuluh *channel vtuber* tersebut adalah sebanyak 4313 video.

3.2. Hasil Data Preprocessing

Setelah *data crawling* dilakukan dan hasilnya disimpan pada file berformat .csv, selanjutnya data memasuki tahap *preprocessing* terlebih dahulu sebelum dilakukannya analisis lebih dalam.

Setelah dikumpulkannya data yang dibutuhkan, didapatkan judul, durasi, deskripsi, dan waktu pengunggahan video, selanjutnya diambil judulnya saja untuk digunakan sebagai input pada tahapan selanjutnya yang menggunakan program yang ada di Zenodo [15] lalu dilakukan proses tokenisasi dengan menggunakan spasi sebagai pemisah kata pada setiap kalimat dalam judul video. Terdapat juga fungsi `lower()` yang berguna untuk mengkonversi semua kata menjadi *lowercase*.

Lalu terdapat pendefinisian *stopword* yang digunakan untuk menyeleksi kata yang kurang informatif serta tidak digunakan dalam penelitian. Di mana beberapa kata yang dimasukkan diseleksi secara manual. Contoh kata yang diseleksi antara lain seperti 2nd, Indonesia, Reine, dan lain sebagainya. Selain itu terdapat juga fungsi `len()` yang berguna untuk menyeleksi karakter special yang tidak diperlukan sehingga kata yang didapatkan sudah bersih dan dapat dijadikan input pada penerapan metode LDA.

3.3. Hasil Penerapan LDA

Setelah selesai memasuki tahap *preprocessing*, data dimasukkan dalam proses penerapan metode LDA. Fungsi *Corpora* berfungsi untuk melakukan pemetaan pada setiap kata dengan menggunakan id kemudian dirubah menjadi *term dictionary*. Kemudian dilakukan *topic modelling* dengan menggunakan modul *genism* yang berguna untuk menghasilkan *document-term matrix*. Adapun *package LdaModel* yang berfungsi untuk mengasalkan model LDA. Sehingga dapat divisualisasikan dengan menggunakan *library Matplotlib* berupa *bar chart*.

Digunakan parameter *perplexity* dan *coherence* untuk mengevaluasi model. *Perplexity* berguna untuk menetapkan nilai optimal dari beberapa topik yang didapatkan. Yang mana semakin rendah nilai *perplexity* maka semakin baik model LDA yang dibuat. Begitu juga sebaliknya, semakin tinggi nilai *coherence* maka semakin bagus model LDA yang dibuat.

3.4. Hasil Analisis

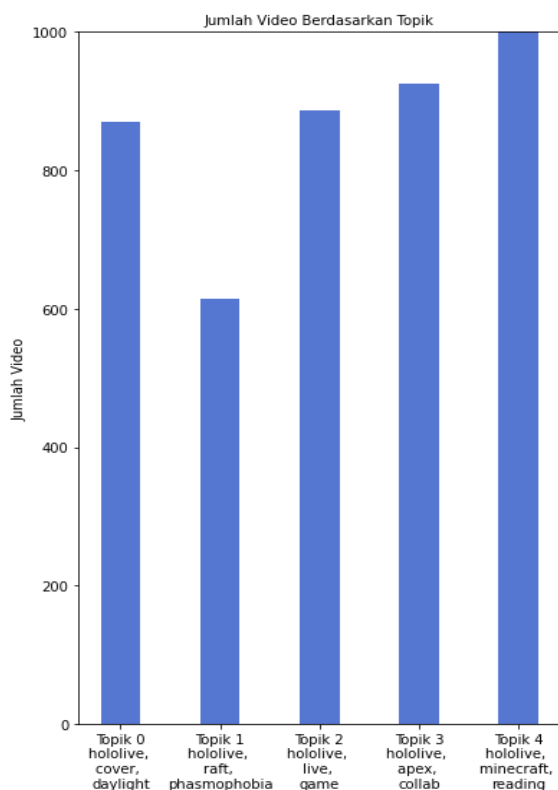
Setelah menerapkan metode LDA pada data, selanjutnya data siap untuk dilakukan analisis. Dalam menentukan jumlah topik yang optimal, dilakukan dengan melihat nilai *perplexity*. Sehingga semakin rendah nilainya maka semakin baik. Begitu pula sebaliknya, jika nilai *topic coherence* semakin tinggi maka semakin bagus modelnya. Dilakukan sebanyak tiga kali percobaan dengan jumlah topik sebanyak tiga,

empat, dan lima. Sehingga didapatkan nilai *perplexity* dan nilai *topic coherence* pada tiga percobaan yang dapat dilihat pada Tabel 3.

TABEL III. NILAI *PERPLEXITY* DAN *TOPIC COHERENCE*

Jumlah Topik	Nilai <i>Perplexity</i>	Nilai <i>Topic Coherence</i>
3	-33.9035688059602	0.5991396465158755
4	-33.97549788856813	0.5595898747436636
5	-34.03315402488048	0.5920112801599459

Dari beberapa percobaan yang telah dilakukan, didapatkan bahwa tidak terdapat perbedaan yang terlalu signifikan berdasarkan banyak topiknya. Namun jika dilihat dari nilai masing-masing *perplexity*, maka dipilih lima topik dengan nilai *perplexity* yang paling rendah. Kemudian dilakukan visualisasi dengan *bar chart* yang menghasilkan topik yang paling koheren yang dapat dilihat pada Gambar 2.



Gambar 2. Hasil dengan Lima Topik

Pada Gambar 2 dapat dilihat *bar chart* yang menunjukkan hubungan antara topik dan jumlah videonya. Dapat dilihat juga pada Tabel 3 yang menunjukkan secara lebih jelas jenis topik beserta jumlahnya. Hasil analisis didapatkan daftar topik yang dominan dari beberapa *vtuber* dengan menggunakan metode LDA.

TABEL IV. HASIL PERSEBARAN SETIAP TOPIK

No	<i>Dominan_Topic</i>	<i>Keywords</i>	Jumlah
0	4.0	hololive, cover, daylight	870
1	4.0	hololive, raft, phasmophobia	615
2	3.0	hololive, live, game	886
3	3.0	hololive, apex, collab	925
4	4.0	hololive, minecraft, reading	1045

Terdapat hasil persebaran topik yang dapat dilihat pada Tabel 3 dan Gambar 2 dalam tiga kata pada setiap topik. Pada topik dominan pertama ditunjukkan hasil yang membahas *member* Hololive yang melakukan *cover* lagu dan juga bermain gim *Dead by Daylight*. Topik dominan pertama ini muncul pada 870 video. Kemudian pada topik dominan kedua ditunjukkan hasil yang membahas *member* Hololive yang bermain gim *Raft* dan *Phasmophobia*. Topik dominan kedua ini merupakan topik yang paling sedikit muncul pada 615 video. Pada topik dominan ketiga ditunjukkan hasil yang membahas *member* Hololive yang sedang melakukan *streaming* dengan konten bermain *game*. Topik dominan ketiga ini muncul pada 886 video. Pada topik dominan keempat ditunjukkan hasil yang membahas *member* Hololive yang sedang berkolaborasi dengan sesama *vtuber* untuk bermain gim *Apex Legends*. Topik dominan keempat ini muncul pada 925 video. Topik dominan kelima ditunjukkan hasil yang membahas *member* Hololive yang bermain gim *Minecraft* sambil membacakan donasi *viewer*. Topik dominan kedua ini merupakan topik yang paling banyak muncul pada 1045 video.

4. KESIMPULAN DAN SARAN

Sudah dilakukan analisis dengan implementasi metode LDA terhadap sepuluh *channel vtuber* Indonesia teratas didapatkan lima topik teratas dari konten video diunggah dan tanyangan langsung. Topik yang dihasilkan dari metode LDA tersebut kebanyakan menyebut Hololive karena *vtuber* teratas tersebut berasal dari agensi *vtuber* asal Jepang dengan nama yang sama. Topik dari konten yang ditayangkan jika diurut maka pertama topik yang populer mengenai gim *minecraft* dilanjutkan dengan *reading* donasi. Kedua merupakan topik gim Apex Legend disertai dengan *collab* atau kolaborasi dengan konten kreator lain, biasanya sesama *vtuber* juga. Ketiga topik mengenai konten *live* atau siaran langsung game secara general tidak spesifik terhadap suatu judul game. Keempat topik mengenai *cover* dari lagu dan gim Dead by Daylight. Kelima merupakan topik yang membahas game *multiplayer* lain seperti Raft dan atau Phasmophobia. Secara keseluruhan, berdasarkan topik yang dihasilkan dari penerapan metode LDA, konten yang ditayangkan *vtuber* kebanyakan tentang topik *video game*, itu juga membuktikan minat penggemar *vtuber* juga pada topik tentang tanyangan *video game*, selain itu tayangan pada hasil analisis juga kebanyakan gim *multiplayer* yang memungkinkan interaksi sesama *vtuber* atau bahkan dengan penonton sangat diminati sehingga makin mendukung bahwa topik mengenai *video game* diminati oleh penggemar *vtuber*.

5. UCAPAN TERIMA KASIH

Ucapan Terima kasih peneliti sampaikan kepada Institut Teknologi Sepuluh Nopember yang telah memberikan dukungan berupa fasilitas selama ini. Terima kasih kepada Ibu Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D selaku pembimbing dari penelitian ini, teman-teman dari peneliti yang selalu memberikan dukungan selama penelitian ini dan seterusnya.

Daftar Pustaka:

- [1] A. Rahmawati, N. L. Nikmah, R. D. A. Perwira, and N. A. Rakhmawati, "Analisis topik konten channel YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation," *Teknologi*, vol. 11, pp. 16–25, Sep. 2021.
- [2] K. Priyadi, Z. Sadiq, and A. Susanti, "ANALISIS MEDIA SIBER PADA SIARAN LANSUNG VIRTUAL YOUTUBER AYUNDA RISU," *MEDIAKOM: Jurnal Ilmu Komunikasi*, vol. 05, 2022.
- [3] P. Andi, "Ketertarikan Mahasiswa Sastra Jepang Universitas Darma Persada Terhadap Perkembangan Virtual Youtuber." 2018. [Online]. Available: <http://repository.unsada.ac.id/id/eprint/765>
- [4] D. R. Puspitaningrum and A. Prasetyo, "Fenomena 'Virtual Youtuber' Kizuna Ai di Kalangan Penggemar Budaya Populer Jepang di Indonesia," *Mediator: Jurnal Komunikasi*, vol. 12, Sep. 2019.
- [5] N. A. Rakhmawati, R. B. Waskitho, D. A. Rahman, and M. F. A. U. Nuha, "Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation," *Journal of Information Engineering and Educational Technology*, vol. 5, pp. 78–83, Sep. 2021.
- [6] H. Puspita Yuri, "Penerimaan Audiens atas Konten Pariwisata dari Virtual Youtuber Andi Adinata," *CoverAge: Journal of Strategic Communication*, vol. 13, no. 1, pp. 38–49, Sep. 2022, doi: 10.35814/coverage.v13i1.3336.
- [7] M. Rossetti, F. Stella, and M. Zanker, "Analyzing user reviews in tourism with topic models," *Information Technology & Tourism*, vol. 16, no. 1, pp. 5–21, Mar. 2016, doi: 10.1007/s40558-015-0035-y.
- [8] I. 'Wayan Dikse Pancane and I. 'Wayan Suriana, "KLASIFIKASI KOMENTAR PUBLIK TERHADAP KEBIJAKAN PEMERINTAH PADA FACEBOOK FRONTPAGE KOMPAS MENGGUNAKAN CLUSTERING K-MEANS, FURTHEST FIRST," *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 3, no. 2, pp. 166–173, Jun. 2020.
- [9] F. Gurcan, O. Ozyurt, and N. E. Cagitay, "Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation," *The International Review of Research in Open and Distributed Learning*, vol. 22, no. 2, pp. 1–18, Jan. 2021, doi: 10.19173/irrodl.v22i2.5358.
- [10] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, "Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm," *J Phys Conf Ser*, vol. 1821, no. 1, p. 012054, Mar. 2021, doi: 10.1088/1742-6596/1821/1/012054.
- [11] L. Mutawalli, M. T. A. Zaen, and W. Bagye, "KLASIFIKASI TEKS SOSIAL MEDIA

- TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 2, no. 2, p. 43, Dec. 2019, doi: 10.36595/jire.v2i2.117.
- [12] P. D. Batlayeri and W. Gatta, "ANALISIS SENTIMEN PEJUALAN JAFRA DALAM PANDEMI COVID-19 DENGAN ALGORITMA KLASIFIKASI," *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, vol. 5, no. 1, pp. 11–18, Apr. 2022.
- [13] A. Erfina and N. Resti Wardani, "ANALISIS SENTIMEN PERGURUAN TINGGI TERMEWAH DI INDONESIA MENURUT ULASAN GOOGLE MAPS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, vol. 5, no. 1, pp. 77–85, Jan. 2022.
- [14] G. B. Herwanto and A. M. Ningtyas, "Recommendation system for web article based on association rules and topic modelling," *Bulletin of Social Informatics Theory and Application*, vol. 1, no. 1, pp. 26–33, Mar. 2017, doi: 10.31763/businta.v1i1.36.
- [15] A. O. Widodo, F. Septiadi, and N. A. Rakhmawati, "Indonesian Vtuber Dataset," Sep. 2022, doi: 10.5281/ZENODO.7128193.