



Tersedia online di www.journal.unipdu.ac.id
Unipdu
Terakreditasi Sinta S5

Halaman jurnal di www.journal.unipdu.ac.id/index.php/teknologi



Analisis topik konten channel YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation

Analysis of Indonesian K-pop YouTube channel content topics using Latent Dirichlet Allocation

Alfrida Rahmawati ^a, Najla Lailin Nikmah ^b, Reynaldi Drajat Ageng Perwira ^c, Nur Aini Rakhmawati ^d

^{a,b,c,d} Sistem Informasi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

email: ^a 0521184000004@mahasiswa.integra.its.ac.id, ^b 05211840000046@mahasiswa.integra.its.ac.id,
^c 05211840000088@mahasiswa.integra.its.ac.id, ^d nur.aini@is.its.ac.id

INFO ARTIKEL

Sejarah artikel:

Menerima 31 Oktober 2020
Revisi 29 Desember 2020
Diterima 3 Januari 2021
Online 23 Januari 2021

Kata kunci:

analisis topik
K-pop
konten YouTube
Latent Dirichlet Allocation
text mining

Keywords:

content YouTube
K-pop
Latent Dirichlet Allocation
text mining
topic analysis

Style APA dalam menyitasi artikel ini:

Rahmawati, A., Nikmah, N. L., Perwira, R. D., & Rakhmawati, N. A. (2021). Analisis topik konten channel YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 11(1), 16-25.

ABSTRAK

Perkembangan teknologi digital telah membawa sebuah media baru, salah satunya YouTube yang kini menjadi salah satu aplikasi yang paling banyak digunakan oleh pengguna internet di dunia. Perkembangan penonton YouTube didukung oleh kontribusi para kreator konten atau disebut juga sebagai YouTuber dari Indonesia. Sehingga semakin banyak tren konten yang diminati oleh penonton Indonesia. Salah satu topik yang kian diminati yakni tentang K-pop. Pada penelitian ini penulis ingin mengetahui topik dominan yang sering diunggah oleh YouTuber K-pop untuk mendukung para kreator konten berkarya. Penelitian ini dilakukan dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Analisis dilakukan setelah melakukan *text mining* pada 2563 video dari 10 akun *youtube* K-pop yang memiliki jumlah *subscriber* lebih dari 100 ribu. Untuk menentukan jumlah topik yang optimal, yakni dengan melihat nilai *perplexity* dan *topic coherence*. Hasil yang didapatkan adalah 5 topik teratas yang menjadi bahan konten dalam video yang diunggah. Topik tersebut antara lain *reaction* untuk *dance cover*, *unboxing* terhadap album dan melakukan *review*, *tebak-tebakan* dari *dance* K-pop dan *vlog* bareng untuk membahas seputar *cover* dan *reaction* terhadap suara pada lagu K-pop.

ABSTRACT

The development of digital technology has brought new media, one of which is Youtube, which is now one of the most widely used applications for internet users in the world. The growth of the audience which is known as viewers, is also supported by the contribution from the content creators or also known as YouTubers from Indonesian. The more the viewers grow, the more their demand for trend content are also growing at surprisingly speed in one of the topics which is H-pop. In this study, the author wanted to see the dominant topics that K-pop YouTubers often upload to support content creator. This research was conducted using the Latent Dirichlet Allocation method. The analysis was carried out on after using text mining on 2563 videos from 10 K-pop YouTuber accounts with more than 100,000 subscribers. To determine the optimal number of topics by looking at the value of perplexity and topic coherence. The results obtained are the top 5 topics that are the content material in the uploaded video. These topics include reactions to dance covers, unboxing on albums and conducting reviews, riddles from K-pop dances and vlogs together to discuss about covers and reactions to sounds on K-pop songs.

Teknologi: Jurnal Ilmiah Sistem Informasi dengan lisensi CC BY NC SA.

1. Pendahuluan

Perkembangan teknologi digital telah membawa sebuah media baru salah satunya yakni YouTube (Cayari, 2011). Platform media yang telah didirikan sejak tahun 2005, kini menjadi salah satu aplikasi

yang paling banyak digunakan oleh pengguna internet di dunia. Tercatat kini ada sebanyak 2 miliar orang yang merupakan pengguna dari YouTube (Iqbal, 2020). Berdasarkan data dari ComScore, ada lebih dari 93 pengunjung dari Indonesia, di mana yang telah berusia di atas 18 tahun untuk menonton video di YouTube (Maulida, 2020).

Tentu saja perkembangan penonton YouTube didukung oleh kontribusi para kreator konten atau disebut juga sebagai YouTuber dari Indonesia (Maulida, 2020). Karena fenomena ini membuat banyak orang menemukan cara baru untuk mengekspresikan diri mereka melalui sarana teknologi untuk mengonsumsi, membuat, dan berbagi sebuah karya video (Cayari, 2011). Sehingga semakin banyak tren konten yang diminati oleh penonton Indonesia diantaranya adalah edukasi, memasak, otomotif, *gaming*, bisnis, dan Korean pop (K-pop) (Maulida, 2020).

Salah satu topik yang kian diminati yakni tentang K-pop, di mana ini didukung oleh adanya fenomena *Korean Wave* (*K-wave*), atau *Hallyu*, di mana ini adalah istilah yang merujuk pada sebuah fenomena budaya Korea mulai dari drama televisi, film, musik, fashion, hingga *game* yang banyak diikuti oleh beberapa negara di Asia (Ahn, Oh, & Kim, 2013). Tentu saja perkembangan pesat ini didukung dengan adanya sosial media seperti YouTube dan Twitter sebagai media persebaran yang membuat K-pop semakin digandrungi oleh khalayak luas.

Dampak yang ditimbulkan oleh adanya fenomena K-pop yang semakin berkembang pesat di Indonesia diantaranya munculnya berbagai produk berbau Korea, rumah makan dengan nuansa Korea dan masih banyak lagi (Juwita, 2018). Karena melihat K-pop sebagai ladang yang bagus untuk berkarya maka juga berdampak pada YouTuber tentang K-pop yang semakin banyak bermunculan. Konten yang dibuat oleh para *youtuber* K-pop cukup diminati oleh banyak penonton, tak jarang konten yang dibuat ditonton hingga satu juta kali. Dengan semakin banyaknya orang yang tertarik dengan konten K-pop, memahami pola penonton dan juga konten yang sering ditonton atau disukai dapat membantu para kreator konten untuk membuat konten yang berkualitas dan lebih menarik banyak penonton. Guna mendukung para kreator konten yang akan terjun di dunia K-pop, maka penting untuk mengetahui topik konten yang diminati oleh penonton.

Berdasarkan uraian di atas, penelitian ini menawarkan solusi untuk melakukan analisis *topic modeling* terhadap konten yang dibuat oleh para *youtuber* K-pop. Penelitian ini akan mengambil data dari YouTuber K-pop dengan jumlah *subscriber* lebih dari seratus ribu ini nantinya akan diperoleh topik konten paling dominan yang diunggah oleh YouTuber K-pop. *Topic modeling* merupakan salah satu pendekatan pada *Text Mining* yang cukup handal dalam melakukan penemuan data-data teks yang tersembunyi dan menemukan hubungan antara teks yang satu dengan lainnya dari suatu *corpus* (Jelodar, et al., 2018).

Penelitian ini dilakukan dengan menggunakan salah satu metode *topic modeling* yaitu metode *Latent Dirichlet Allocation* (LDA) yang digunakan untuk membantu menentukan topik yang muncul pada akun YouTuber K-pop. LDA ini dilakukan dengan meringkas, melakukan klasterisasi, menghubungkan hingga memproses data agar menghasilkan daftar topik yang diberi bobot untuk setiap dokumen (Campbell, Hindle, & Stroulia, 2014). Selain itu, LDA ini merupakan salah satu metode *topic modelling* yang sedang populer saat ini. Banyak penelitian yang menggunakan metode LDA ini sehingga menjadi salah satu alasan untuk menggunakan teknik tersebut.

Sehingga diharapkan penelitian ini dapat memberikan manfaat bagi kreator konten agar mengetahui topik dominan yang sering diunggah oleh YouTuber K-pop, sehingga dapat dijadikan referensi mereka untuk berkarya. Selain itu praktisi akademik dapat menjadikan karya tulis ilmiah ini sebagai bahan rujukan untuk penelitian yang berkaitan.

Penelitian ini memiliki referensi dari penelitian Rangkuti (2020), perbedaannya dengan penelitian ini adalah penggunaan studi kasus YouTuber K-pop. Selain itu, sebelumnya menggunakan komentar sebagai sumber data, sedangkan pada penelitian ini menggunakan data utama yang diambil salah satunya adalah judul video untuk menentukan topik. Dengan menggunakan metode yang sama yakni LDA, penelitian ini diharapkan bisa memberikan gambaran bagi para kreator YouTube sebagai referensi mereka untuk berkarya.

2. State of the Art

2.1. Korean pop

Korean Wave (K-wave) atau *Hallyu* merupakan istilah yang merujuk pada budaya Korea yang tersebar secara global melalui drama televisi, film, lagu pop, dan selebriti terkait yang telah mendapatkan popularitas yang luar biasa di China, Taiwan, Hong Kong, negara-negara Asia Timur dan Tenggara lainnya (Shim, 2006). Tak terkecuali di Indonesia, di mana fenomena *K-wave* ini digencarkan melalui drama Korea hingga *music pop* Korea (K-pop). K-Pop adalah musik modern dari Korea. Musik Korea ini mampu memenuhi selera musik dunia khususnya Amerika Serikat, sehingga musik Korea mampu mendominasi pasar *music* dunia. Meski ada juga J-pop (musik pop jepang) dan aliran musik lainnya. Musik pop Korea lebih mampu memuaskan selera generasi sekarang, sehingga K-Pop lebih digemari. Masyarakat Korea telah berhasil menanamkan pentingnya musik dalam kehidupan sehari-hari. Akibatnya, konsumsi musik di Korea sangat tinggi dalam hal musik tradisional, pop, *rock*, *jazz*, atau klasik. modern K-pop muncul pada maret 1992, ketika Seo Taiji and *Boys* merilis album pertama mereka kemudian diikuti oleh banyak produser dan grup baru antara 1994 dan 1997 yang menyebabkan K-pop menjadi lebih berkembang.

2.2. *Youtuber K-pop*

Youtube adalah platform *streaming* video terbesar dan paling dikenal di internet. situs web ini memiliki jutaan pengguna aktif harian dari seluruh dunia dan menampung miliaran video. Dunia hiburan dan musik pop Korea (biasa disebut sebagai K-Pop) dengan cepat menjadi salah satu pasar terbesar, dengan video yang ditonton miliaran dalam beberapa tahun terakhir. Dikutip *Korean Culture Informative Service*, jumlah orang yang menonton K-Pop melalui Youtube di Asia adalah 566.273,899 (KOCIS, 2011). Karena industri K-pop terus tumbuh dan berkembang secara internasional, semakin banyak saluran yang mengunggah video dengan konten K-pop tidak terkecuali di Indonesia.

Youtuber adalah seseorang yang mengunggah, memproduksi, atau tampil di video yang ada di platform untuk berbagi video yaitu YouTube. Karena popularitas musik K-pop yang terus berkembang di Indonesia dan diterima dengan baik oleh masyarakat, hal ini menumbuhkan kreativitas para *fans* K-pop atau yang biasa disebut *K-popers* untuk memproduksi konten tentang *music pop* Korea pun terus berkembang pula. Sehingga mengumpulkan penggemar dari berbagai kalangan dari anak kecil hingga dewasa dengan menjadi *subscriber* dari saluran kreator tersebut. *Subscriber* merupakan seseorang yang berlangganan terhadap sebuah saluran YouTube dan dapat mendapatkan informasi terbaru mengenai kanal tersebut.

2.3. *Text mining*

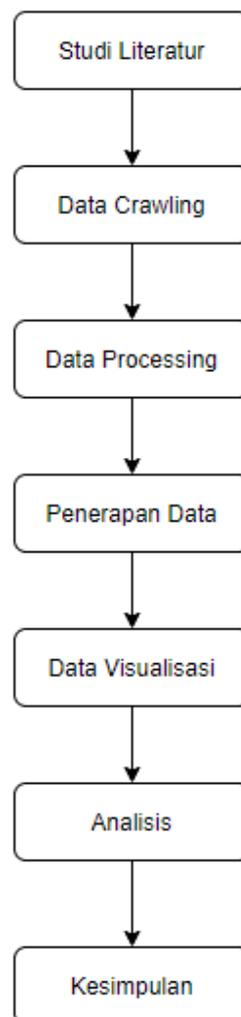
Text mining merupakan salah satu bidang yang luas saat ini untuk menganalisis dokumen dengan bahasa natural. *Text mining* dapat didefinisikan sebagai suatu proses menggali informasi di mana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tool* analisis yang merupakan komponen-komponen dalam *data mining* (Brata & Hetami, 2015). *Text mining* digunakan untuk mengolah dokumen sebelum dilakukan proses *similarity*. *Text mining* merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. *Text mining* memiliki tujuan dan menggunakan proses yang sama dengan *data mining*, tetapi memiliki masukan yang berbeda. Masukan untuk *Text mining* adalah data yang tidak (atau kurang) terstruktur, seperti dokumen *Word*, PDF, dan kutipan teks, sedangkan masukan untuk *data mining* adalah data yang terstruktur (Siregar, Sinaga, & Arianto, 2017). Bidang ini mungkin lebih dikenali sebagai proses dalam menganalisis teks untuk mengekstrak informasi yang berguna untuk suatu tujuan tertentu. Dibandingkan dengan jenis data yang tersimpan dalam *database*, *text mining* menggunakan data teks yang tidak terstruktur, tidak memiliki bentuk yang jelas dan sulit untuk diuraikan dengan pendekatan algoritma. Namun, dalam budaya modern, teks adalah perantara yang paling umum untuk pertukaran secara formal dari informasi.

Text mining dalam prakteknya mencari pola-pola tertentu, mengasosiasikan satu bagian teks dengan lain berdasar aturan-aturan tertentu, kata-kata yang dapat mewakili sehingga dapat dilakukan analisis keterhubungan antar satu dengan lain, dalam 7 kumpulan dokumen yang sangat banyak. Dokumen yang ada bisa bersifat statis, yaitu dokumen yang tidak akan diperbarui lagi ataupun dinamis yaitu dokumen yang akan selalu diperbarui dalam rentang waktu tertentu.

2.4. *Metode Latent Dirichlet Allocation (LDA)*

Latent Dirichlet Allocation (LDA) merupakan metode *topic modeling* dan topik analisis yang paling populer saat ini. LDA muncul sebagai salah satu metode yang dipilih dalam melakukan analisis pada dokumen yang berukuran sangat besar. LDA dapat digunakan untuk meringkas, melakukan klusterisasi, menghubungkan maupun memproses data yang sangat besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen (Campbell, Hindle, & Stroulia, 2014). LDA mengidentifikasi informasi topik tersembunyi dalam koleksi dokumen besar menggunakan pendekatan *bag of words* (cara representasi data teks) yang memperlakukan setiap dokumen sebagai vektor jumlah kata dan direpresentasikan sebagai distribusi probabilitas atas beberapa topik, sementara setiap topik direpresentasikan sebagai distribusi probabilitas atas sejumlah kata (Hong & Davison, 2010). Mekanisme kerja LDA terbagi menjadi dua bagian yaitu penalaran dan implementasi. Penalaran merupakan proses LDA untuk menentukan bobot dari setiap kata yang ada pada setiap dokumen dalam korpus. Implementasi merupakan tahap penerapan LDA untuk kebutuhan temu kembali informasi selanjutnya (Prihatini, Suryawan, & Mandia, 2017).

3. Metode Penelitian



Gambar 1. Alur Penelitian

3.1. Alur penelitian

Pada penelitian ini, akan dilakukan serangkaian proses untuk mendapatkan hasil akhir sesuai tujuan yang dicantumkan. Terdapat 7 proses yang akan dilakukan pada penelitian ini yang terdiri dari studi literatur, *data crawling*, *data processing*, penerapan LDA, data visualisasi, analisis, dan kesimpulan. Adapun metode yang digunakan dalam penelitian ini yakni metode LDA. Data yang digunakan dalam penelitian ini merupakan judul konten K-pop dari *website* YouTube. Adapun Gambar 1 merupakan diagram yang menunjukkan alur penelitian ini dari awal hingga akhir.

3.2. *Studi literatur*

Hal pertama yang dilakukan sebagai langkah awal dalam penelitian ini adalah penulis melakukan studi literatur dan juga penggalian informasi terkait dengan metode dan cara kerja dari jurnal terdahulu yang berkaitan dengan konten penulisan ini sebagai referensi dalam penelitian. Studi literatur dilakukan dengan merujuk pada jurnal, artikel ilmiah, skripsi dan *website* resmi terpercaya yang dapat mendukung berjalannya penulisan ilmiah kali ini. Beberapa informasi yang didapatkan meliputi penjelasan mengenai Youtube, K-pop, pengertian LDA hingga metode penelitian.

3.3. *Crawling data*

Crawling data dilakukan untuk mengumpulkan data pada sebuah *website* yakni YouTube. URL YouTube akan menjadi acuan untuk mencari semua *hyperlink* yang kemudian akan dilakukan *indexing* untuk mencari data dalam dokumen pada setiap *link* ada. Data yang akan dicari adalah data yang terkait dengan topik K-pop. Topik ini nantinya akan menargetkan beberapa *channel* YouTube yang kontennya sebagian besar terisi oleh konten K-pop. *Crawling data* kemudian dilanjutkan dengan menghimpun terlebih dahulu 10 akun YouTuber K-pop yang nantinya dijadikan target *crawling*. Selanjutnya akan dilakukan penambangan data untuk mendapatkan data berupa judul, jumlah *subscriber*, dan jumlah video yang akan dihimpun untuk proses penelitian berikutnya.

Pada awalnya penelitian ini membutuhkan API *key* dan *channel id* dari masing-masing akun YouTuber K-pop sebagai bahan untuk melakukan *scraping*. Di mana *scraping* merupakan teknik untuk mengumpulkan data dari sebuah situs secara otomatis. Rangkuti (2020) dalam penelitiannya memanfaatkan *website* commentpicker.com untuk mengetahui setiap *channel id* dari masing-masing YouTuber K-pop.

3.4. *Data preprocessing*

Data yang telah didapatkan kemudian dilakukan proses *cleaning*, sehingga data menjadi bentuk yang lebih teragregasi sehingga dapat mempermudah tahap pemrosesan data. Setiap data yang sejenis akan dimasukkan ke dalam suatu *index* yang bertujuan untuk membedakan dokumen satu dengan dokumen lainnya.

Dari sekumpulan data yang telah didapatkan kemudian dilakukan proses tokenisasi yang bertujuan untuk memisahkan teks ke dalam unit-unit kecil yang biasa disebut dengan token (Rangkuti, 2020).

3.5. *Penerapan LDA*

Data yang telah teragregasi dalam beberapa *index* kemudian diterapkan metode LDA. Penerapan LDA bertujuan untuk memperoleh persebaran data yang dapat membentuk suatu proyeksi linear yang akan menghasilkan topik untuk suatu dokumen. Penerapan LDA akan dibagi menjadi beberapa tahap. Tahap yang pertama yakni mengisi parameter yang akan digunakan sebagai batas data yang digunakan. Selanjutnya adalah menerapkan semi *random distribution*, yang mana didasari atas metode distribusi Dirichlet. Setelah itu dilakukan iterasi untuk melihat parameter yang dapat menentukan distribusi dari topik dan kata dari dokumen yang diharapkan.

Untuk menentukan banyaknya topik secara optimal dapat dilakukan dengan menggunakan perhitungan *perplexity*. Di mana nilai *perplexity* itu didapatkan dari perhitungan dengan menentukan kemungkinan dari log teks dokumen yang tidak terlihat. Jika hasil yang didapatkan nilai *perplexity* rendah, maka model tersebut adalah baik. Terkadang cara dengan menggunakan *perplexity* ini tidak berkorelasi terhadap interpretasi manusia dengan baik (Rangkuti, 2020).

3.6. *Data visualisasi*

Data yang didapatkan dari penerapan LDA akan menjadi data mentah yang kemudian dilakukan visualisasi dengan memanfaatkan grafik dan *bar-chart* untuk mempermudah proses analisis data berikutnya. Data visualisasi ini akan dilakukan dengan menunjukkan seberapa besar perbandingan persebaran data yang satu dengan data yang lainnya.

3.7. *Analisis*

Input dari proses analisis ini adalah *raw data* yang telah didapatkan sebelumnya menggunakan metode LDA beserta visualisasi data yang telah dilakukan pada tahap sebelumnya. Pada proses analisis ini,

data akan dianalisis menggunakan metode *topic modelling* yang dapat mengelompokkan topik-topik tertentu berdasarkan kemiripan pola atau kesamaan kata kunci yang ditemukan. Hasil visualisasi data sebelumnya yang berbentuk grafik juga akan dianalisis terkait proporsi topik yang ditemukan. Hasil dari analisis ini akan terus dievaluasi untuk menemukan topik-topik yang dominan dalam suatu konten dan konten-konten yang dominan dalam setiap *channel* (Budiman, Santoso, & Afirianto, 2017).

3.8. Kesimpulan

Data yang telah dianalisis akan menghasilkan kesimpulan dari 10 (sepuluh) *channel* yang telah dijadikan subjek penelitian ini. Kesimpulan yang didapatkan akan memuat empat topik teratas yang menjadi pokok bahasan utama dalam konten-konten yang diunggah pada *website* YouTube tersebut. Kesimpulan juga akan memuat setidaknya empat topik K-pop yang dominan dikalangan penikmat K-pop masa kini.

4. Hasil dan Pembahasan

4.1. Hasil crawling data

Hal pertama yang dilakukan yakni penambangan data. Di mana data yang digunakan dalam penelitian ini adalah data akun mulai dari jumlah *subscriber*, nama saluran, hingga judul dan deskripsi video dari beberapa YouTuber K-pop. Data yang dikumpulkan yakni data keseluruhan mulai dari awal masing-masing saluran itu terbentuk hingga terakhir tanggal 21 Oktober 2020.

Pengambilan data dilakukan terhadap YouTuber K-pop yang telah memenuhi kriteria tertentu. Di mana kriteria tersebut adalah YouTuber yang kerap mengunggah konten mengenai K-Pop dan telah memiliki lebih dari seratus ribu *subscriber*. Dalam pemilihan YouTuber tersebut dilakukan secara acak dengan memperhatikan kriteria yang ada. Berdasarkan kriteria tersebut, maka diambil sepuluh sampel YouTuber K-pop sebagai bahan penelitian. Daftar nama saluran yang diambil ditunjukkan pada Tabel 1. yang juga menampilkan jumlah *subscriber* dari akun tersebut. Hasil yang diperoleh akan dimanfaatkan sebagai masukan dalam metode LDA pada tahap selanjutnya.

Tabel 1. Daftar akun YouTuber dan jumlah *subscriber*

Nama Saluran	Jumlah Subscriber (k=ribu)
agorivalll	372k
Akmal Ramdan	228k
Alphiandi	707k
Anthony Yaputra	322k
Geraldyan Channel	193k
Kevin Ferdiansyah	525k
NADAFID	395k
Natya Shina	525k
Tiffani Afifa	493k
Yunitafauziah CHIPMUNK	536k

Tabel 2. Daftar jumlah video dari YouTuber

Nama Saluran	Jumlah Video
agorivalll	73
Akmal Ramdan	558
Alphiandi	422
Anthony Yaputra	169
Geraldyan Channel	89
Kevin Ferdiansyah	451
NADAFID	191
Natya Shina	271
Tiffani Afifa	256
Yunitafauziah CHIPMUNK	83
Total Video	2563

Pada Tabel 1 dapat dilihat bahwa *subscriber* dari YouTuber K-pop itu memiliki jumlah yang beragam. Mulai dengan jumlah yang paling sedikit, yakni sejumlah 193 ribu *subscriber* dan dimiliki oleh Geraldtyan Channel. Sedangkan jumlah *subscriber* paling banyak dimiliki oleh Alphiandi dengan jumlah 707 ribu.

Setelah API *key* dan *channel id* diperoleh, maka didapatkan judul serta deskripsi yang berisi waktu pengunggahan video dan keterangan dari setiap akun YouTuber tersebut. Setelah data judul, durasi, deskripsi dan waktu pengunggahan video itu didapatkan kemudian dihimpun dalam bentuk tabel yang disimpan dalam *file* berekstensi *.csv* sebagai bahan masukkan proses analisis selanjutnya dan diunggah melalui akun Zenodo (Nikmah, Rahmawati, Perwira, & Rakhmawati, 2020).

Pada Tabel 2 diperoleh jumlah video dari masing-masing akun YouTuber K-pop. Hasil yang didapatkan menunjukkan setiap saluran memiliki jumlah video yang beragam, mulai dari 70 hingga mencapai 500 video yang telah diunggah. Jika keseluruhan video dari sepuluh saluran di atas itu dijumlahkan, maka terdapat sebanyak 2563 judul video yang telah dibuat oleh kreator tersebut.

4.2. Hasil Data Preprocessing

Setelah data dikumpulkan dengan cara *scraping*, kemudian hasilnya yang dihimpun masing-masing dalam bentuk *.csv*. Kemudian untuk mendapatkan hasil yang maksimal, sebelumnya data yang diperoleh tersebut dilakukan *preprocessing* terlebih dahulu sebelum dilakukan analisis lebih lanjut.

Dari sekumpulan data yang telah didapatkan, di mana berupa judul, durasi, deskripsi dan waktu pengunggahan video kemudian dipilihlah judul sebagai masukan untuk analisis. Dengan menggunakan program yang ada di Zenodo (Nikmah, Rahmawati, Perwira, & Rakhmawati, 2020) ini kemudian dilakukan proses tokenisasi dengan memanfaatkan spasi sebagai pemisah kata dalam sebuah kalimat judul video. Adanya fungsi *lower()* ini untuk mengubah karakter yang berbentuk *uppercase* menjadi bentuk *lowercase*.

Kemudian dilakukan tahapan *stopword* untuk menghilangkan beberapa kata yang tidak informatif dan tidak diperlukan dalam penelitian. Kata-kata yang dimasukkan dalam *stopword* ini didefinisikan secara manual. Contohnya saja seperti *di*, *ke*, nama *youtuber* yang tercantum dalam judul, dan lain-lain. Selain itu terdapat fungsi *len()* untuk tidak mengikut sertakan tanda baca yang tidak perlu. Sehingga data yang didapatkan berupa kata yang bisa dimasukkan dalam proses LDA.

4.3. Hasil Penerapan LDA

Data yang telah dilakukan *preprocessing* kemudian masuk dalam proses LDA. Dengan menggunakan fungsi *Corpora* untuk melakukan pemetaan pada setiap kata dengan menggunakan *id*, data yang ada sebelumnya itu kemudian diubah menjadi *term dictionary*. Lalu dilakukan *topic modelling* dengan menggunakan modul *gensim* untuk menghasilkan *document-term matrix*. Untuk menghasilkan LDA model maka menggunakan *package LdaModel* dan kemudian menggunakan *CoherenceModel* untuk menghitung banyaknya topik yang saling koheren. Di mana semakin tinggi nilai *topic coherence* maka semakin bagus hasilnya. Untuk menetapkan nilai optimal dari banyaknya topik dapat dilakukan dengan menggunakan perhitungan *perplexity*. Di mana model LDA yang baik itu dilihat dari nilai *perplexity* yang rendah.

4.4. Hasil Analisis

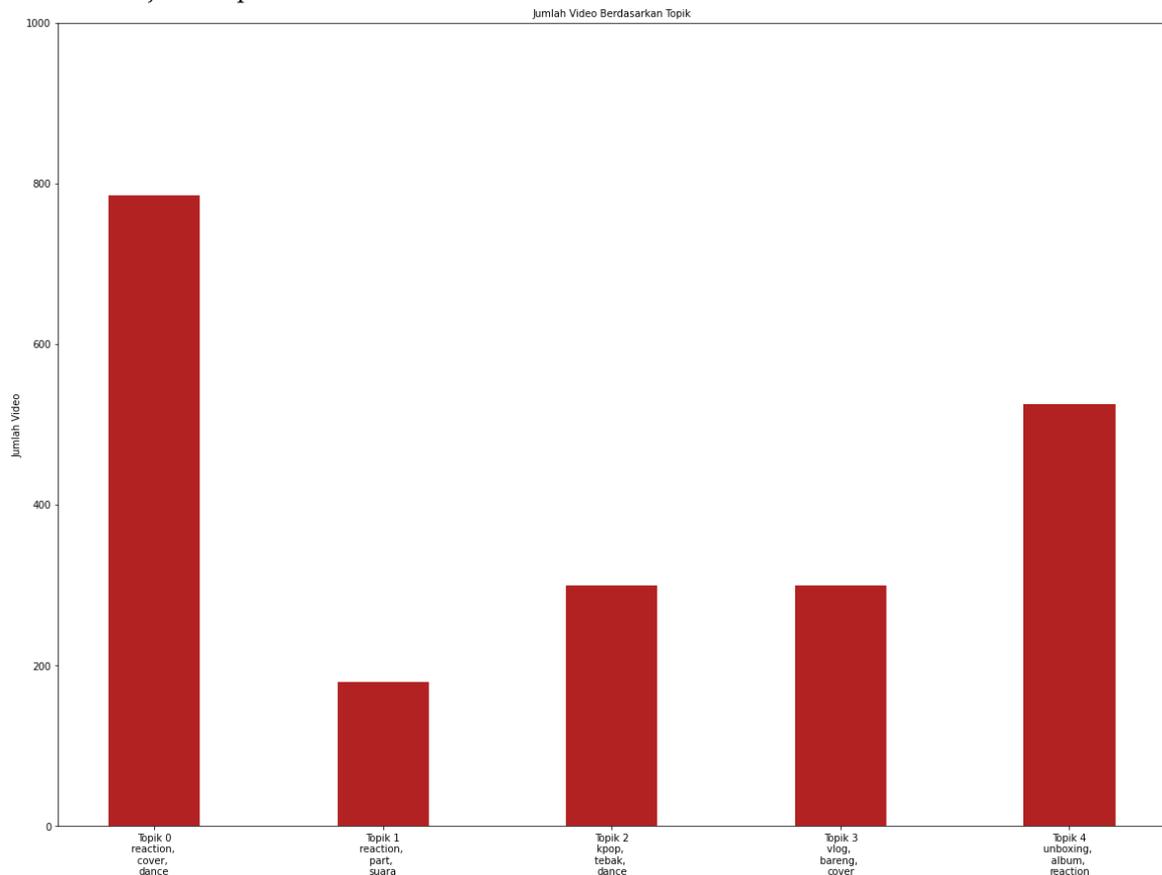
Setelah menerapkan LDA pada data, kemudian data tersebut dianalisis. Untuk menentukan jumlah topik yang optimal yakni dengan melihat nilai *perplexity*, semakin rendah nilainya semakin baik. Sebaliknya jika nilai *topic coherence* semakin tinggi semakin bagus. Dengan mencoba melakukan perhitungan dengan jumlah topik sebanyak 3, 4 dan 5 didapatkan nilai *perplexity* dan *topic coherence* yang ditampilkan dalam Tabel 3.

Tabel 3. Nilai *perplexity* dan *topic coherence*

Jumlah Topik	Nilai <i>Perplexity</i>	Nilai <i>Topic Coherence</i>
3 Topik	-9,336946373684114	0,6211897616146652
4 Topik	-9,349112863478688	0,6014899823663401
5 Topik	-9,36608474109665	0,6294236743113695

Dari hasil yang didapatkan menunjukkan bahwa nilai yang ada tidak menunjukkan perbedaan yang jauh signifikan. Jika melihat dari nilai *perplexity* untuk menentukan jumlah topik yang baik maka

dipilihlah dengan banyak 5 topik, di mana menunjukkan nilai *perplexity* paling rendah. Selain itu pada 5 topik juga memiliki nilai *topic coherence* memiliki nilai tertinggi. Kemudian hasil topik yang paling koheren ditunjukkan pada Gambar 2.



Gambar 2. Hasil dengan 5 topik

Di mana dapat dijabarkan secara lebih jelas detail melalui Tabel 4 yang menunjukkan jenis topik beserta jumlahnya. Pada analisis data sehingga dihasilkan daftar topik yang dominan dari para YouTuber K-pop dengan menggunakan metode LDA.

Tabel 4. Hasil persebaran setiap topik

No	Dominan_Topic	Keywords	Jumlah
1	0,0	reaction, cover, dance	785
2	0,1	reaction, part, suara	179
3	0,2	kpop, tebak, dance	299
4	0,3	vlog, bareng, cover	299
5	0,4	unboxing, album, reaction	525

Dari Tabel 4 dan Gambar 2 menunjukkan hasil persebaran topik ditampilkan dalam 3 kata di setiap topiknya. Pada topik 0 menunjukkan hasil yang membahas mengenai *reaction* seputar *dance cover*. Topik 0 ini muncul pada video paling banyak yakni dengan jumlah 785 video. Kemudian, pada topik 1 membahas seputar *reaction* yang berhubungan dengan suara orang pada *part* atau bagian dari sebuah lagu K-pop. Topik satu ini muncul pada 179 video yang ada. Selanjutnya, terdapat topik 2 yang membahas mengenai tebak-tebakan dari *dance* K-pop. Topik 2 ini muncul pada sebanyak 299 video. Kemudian pada topik 3 menunjukkan hasil di mana YouTuber itu melakukan *vlog* bareng, biasanya mereka berkolaborasi dengan YouTuber lain untuk membahas seputar *cover*. Topik 3 kali ini muncul pada 299 video. Terakhir yakni topik 4 yang membahas *reaction* tentang *unboxing album*. Di mana YouTuber melakukan *unboxing* terhadap album dan melakukan *review*. Topik ini memiliki jumlah yang terbanyak kedua dengan 525 video.

5. Kesimpulan

Berdasarkan analisis dari metode LDA yang dilakukan pada sepuluh saluran YouTuber K-pop, didapatkan 5 topik teratas yang menjadi bahan konten dalam video yang diunggah. Topik tersebut

antara lain *reaction* untuk *dance cover* dan suara pada *part* dari sebuah lagu K-pop hingga tebak-tebakan dari *dance* K-pop. Selanjutnya ada *vlog* bareng dan *unboxing album*. Jika diurutkan dari topik paling dominan, posisi pertama adalah topik tentang *reaction* untuk *dance cover*. Kemudian yang kedua adalah *unboxing* terhadap album dan melakukan *review*. Selanjutnya di urutan ketiga dan keempat adalah tebak-tebakan dari *dance* K-pop dan *vlog* bareng untuk membahas seputar *cover*. Urutan kelima terdapat *reaction* terhadap suara pada *part* dari sebuah lagu K-pop. Dari hasil penelitian ini menunjukkan bahwa mayoritas topik-topik tersebut merupakan topik yang relevan dengan konten YouTuber K-pop di Indonesia.

6. Referensi

- Ahn, J., Oh, S., & Kim, H. (2013, Juli 1). Korean Pop Takes Off! Social Media Strategy of Korean Entertainment Industry. *10th International Conference on Service Systems and Service Management* (pp. 774-777). Hong Kong: IEEE.
- Brata, D. W., & Hetami, A. (2015). Perancangan Information Retrieval (IR) untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris dengan Pembobotan Vector Space Model. *Jurnal Ilmiah Te. Jurnal Ilmiah Teknologi dan Informasi ASIA*, 9(1), 53-59.
- Budiman, E., Santoso, E., & Afirianto, T. (2017). Pendeteksi Jenis Autis pada Anak Usia Dini Menggunakan Metode Linear Discriminant Analysis (LDA). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1, 583-592. Retrieved Oktober 21, 2020, from <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/169/85>
- Campbell, J. C., Hindle, A., & Stroulia, E. (2014). *Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data*.
- Cayari, C. (2011). The YouTube effect: How YouTube has provided new ways to consume, create, and share music. *International Journal of Education & the Arts*, 12(6).
- Hong, L., & Davison, B. D. (2010). Empirical Study of Topic Modeling in Twitter. *SOMA '10: Proceedings of the First Workshop on Social Media Analytics*, 10, 80-88.
- Iqbal, M. (2020, Oktober 15). *YouTube Revenue and Usage Statistics (2020)*. Retrieved from Business of Apps: <https://www.businessofapps.com/data/youtube-statistics/>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2018). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimed Tools Appl*, 78, 15169–15211.
- Juwita, S. H. (2018). Tingkat Fanatisme Penggemar K-Pop dan Kemampuan Mengelola Emosi pada Komunitas Exo-L di Kota Yogyakarta. *Jurnal Riset Mahasiswa Bimbingan dan Konseling*, 4(7), 273-286. Retrieved Oktober 21, 2020, from <http://journal.student.uny.ac.id/ojs/index.php/fipbk/article/viewFile/12568/12114#:~:text=Menurut%20hasil%20penelitian%20yang%20dilakukan,dan%20%20%25%20berusia%20di%20atas>
- KOCIS. (2011). *The Korean Wave: A New Pop Culture Phenomenon*. South Korea: Korean Culture and Information Service.
- Maulida, L. (2020, September 15). *Jumlah pengguna unik YouTube di Indonesia capai 93 juta*. Retrieved Oktober 21, 2020, from tek.id: <https://www.tek.id/tek/jumlah-pengguna-unik-youtube-di-indonesia-capai-93-juta-b1ZT79iPE>
- Nikmah, N. L., Rahmawati, A., Perwira, R. D., & Rakhmawati, N. A. (2020). Dataset from YouTuber K-pop in Indonesia. Retrieved from <https://zenodo.org/record/4153459#.YAr3FOgzbIX>
- Prihatini, P. M., Suryawan, I. K., & Mandia, I. N. (2017). Metode Latent Dirichlet Allocation untuk Ekstraksi Topik Dokumen. *Jurnal Logic*, 17(3), 154-158.
- Rangkuti, M. D. (2020). *Analisis Topik Komentar Video Beberapa Akun Youtube e-Commerce Indonesia Menggunakan Metode Latent Dirichlet Allocation*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah. Retrieved Oktober 19, 2020, from <http://repository.uinjkt.ac.id/dspace/bitstream/123456789/51580/1/MUHAMMAD%20DIAZ%20ELPRANA%20RANGKUTI-FST.pdf>
- Shim, D. (2006). Hybridity and the rise of Korean popular culture in Asia. *Media, Culture & Society*, 28(1), 25-44.

Siregar, R. R., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Computatio: Journal of Computer Science and Information Systems*, 1(2), 171-186.