

SEGMENTASI POPULARITAS AKUN YOUTUBE MENGGUNAKAN METODE ID3

SEGMENTATION OF YOUTUBE ACCOUNT POPULARITY USING ID3 METHOD

Bernadus Gunawan Sudarsono^{1)*}, Rahma Saputra²⁾, Ferik Utomo³⁾, dan Christian Wijaya⁴⁾

¹⁾Program Studi Sistem Informasi, Universitas Bung Karno, Jakarta
^{2,3,4)}Program Studi Sistem Informasi, Universitas Bunda Mulia, Jakarta

Diterima 09 Agustus 2020 / Disetujui 14 Agustus 2020

ABSTRACT

Internet isn't a new thing at this day, because almost every people in different background in around of the world can access internet so easily. With easy internet access, any people can using social media everyday and anytime whoever they wants, one of the social media example is Youtube, some platform in internet that provide internet-based shows or videos. Youtube was founded in 2005, but has only been popular and accessible in recent years. On YouTube there is data about activities on YouTube itself, mainly data about YouTube channels that have a number of subscribers, total views, and others. In this journal we will discuss the grade obtained by a YouTube channel by processing Big Data about the YouTube channel. The research method used is a descriptive research method where data will be processed to make a systematic, factual, and accurate description of the facts or the nature of a particular population. The research aims to examine the popularity level of several YouTube channels from the number of uploads and subscribers and also the number of people who view the video. The results showed that the level of popularity (which was assessed based on view, subscriber, and upload) of a total of 5001 different youtube channels can be in the following order: (A ++, A +, A, A-, B +).

Keywords: Big Data, Social Media, Youtube, Grade, Popularity.

ABSTRAK

Internet bukan merupakan hal yang awam di era digital seperti ini, hampir disetiap kalangan dunia dapat mengakses internet dengan mudah. Dengan akses internet yang mudah didapatkan hampir seluruh kalangan masyarakat di Indonesia, membuat masyarakat mudah mengakses media sosial yang tersedia didalam internet. Salah satu media yang menyediakan tontonan atau video dengan berbasis internet adalah Youtube. Youtube berdiri sejak tahun 2005, namun baru populer dan dapat diakses oleh masyarakat Indonesia secara luas dalam beberapa tahun terakhir. Didalam youtube terdapat data-data mengenai aktivitas didalam youtube itu sendiri, terutama adalah data mengenai channel youtube yang memiliki jumlah subscriber, total views, dan lain-lain. Didalam jurnal ini kami akan membahas mengenai grade yang didapatkan oleh suatu channel youtube dengan dilakukannya pengolahan Big Data mengenai channel youtube. Metode penelitian yang dilakukan adalah metode penelitian deskriptif dimana data akan diolah untuk membuat deskripsi secara sistematis, faktual, serta akurat pada fakta atau sifat populasi tertentu. Penelitian bertujuan untuk menguji tingkat popularitas beberapa channel youtube dari banyaknya upload serta pelanggan dan juga banyaknya orang yang melihat video. Hasil penelitian menunjukkan bahwa adanya tingkat popularitas (yang dinilai berdasarkan view, subscriber, dan upload) dari total 5001 channel youtube yang berbeda dapat yang mempunyai urutan sebagai berikut : (A++, A+, A, A-, B+).

Kata Kunci: Big data, Media sosial, Youtube, Grade, Popularitas.

*Korespondensi Penulis:

E-mail: gunawanbernadus@ubk.ac.id

PENDAHULUAN

Pengertian dari data itu adalah, “Suatu bahan mentah bagi informasi, yang dirumuskan sebagai kelompok lambang-lambang tidak acak yang menunjukkan jumlah-jumlah, tindakan-tindakan, hal-hal, dan sebagainya. Data- data disusun untuk mengolah tujuan-tujuan menjadi susunan data, susunan kearsipan, & pusat data / landasan data (Davis, 2012). Sedangkan pengertian dari *Big Data* itu sendiri adalah “Data yang melebihi suatu proses kapasitas dari kovensi sistem *database* yang sudah ada. Data yang terlalu besar dan terlalu cepat ataupun yang tidak sesuai dengan struktur aritektur *database* yang sudah ada” (Dumbill, 2012). Maka dapat disimpulkan bahwa *Big Data* adalah suatu istilah yang menggambarkan *volume* suatu data yang besar secara terstruktur maupun tidak terstruktur.

Big Data ini sudah diterapkan diberbagai macam jenis kegiatan, contohnya bisnis, organisasi, dan contoh yang ingin dibahas ini adalah media sosial / *Youtube*. Tujuan pengolahan *Big Data* adalah agar dapat menghasilkan suatu informasi ataupun wawasan yang sebelumnya belum dimiliki, dengan tujuan akhir sebagai bantuan untuk menganalisis keputusan keputusan yang ingin diambil, agar keputusan yang dipilih sesuai dengan keinginan atau kebutuhan.

Di dalam *Big Data* juga terdapat tiga bagian penting diantaranya adalah *Volume*, *Variety*, dan *Velocity*. *Volume*, menyangkut tentang besarnya ukuran dari *Big Data*, terutama yang terdapat pada “*Youtube*”, mengenai *channel youtube* beserta dengan variabel-variabel lainnya dengan contoh *subscriber*, *watch time*, dst. *Volume* dari data ini juga akan terus bertambah disetiap waktunya, maka data yang tersimpan akan sangat banyak. *Variety*, dengan artian *Big Data* memiliki variasi atau jenis data yang beragam didalamnya, dengan contoh audio, pesan, teks, dan lain-lain. *Velocity*, memiliki arti yaitu kecepatan, namun didalam *Big Data*

yaitu kecepatan yang mereferensi pada peningkatan pengumpulan data dan seberapa cepat data yang dikumpulkan harus dapat dianalisa ulang agar informasi yang diterima dapat berguna.

Pengertian sosial media sendiri adalah sebuah media yang bersifat *online* dimana para penggunanya dapat berinteraksi, berbagi, dan menciptakan isi didalamnya berupa blog, wiki, jejaring sosial, *forum*, dan dunia *virtual* (Cahyono, 2016). Bentuk media sosial yang sering digunakan pada umumnya adalah *blog*, *wiki*, dan *forum*, sedangkan *Youtube* adalah media sosial yang berisikan video didalamnya. Pada saat ini *Youtube* juga menjadi media sosial yang sangat dominan di dunia. Diperkirakan 20 jam durasi video yang diupload ke *youtube*, setiap menitnya dengan 6 miliar views perhari (Fatty, dkk., 2016). Banyak hal yang membuat *youtube* banyak diminati oleh para penggunanya, diantaranya adalah fitur-fitur yang ditawarkan beserta konten-konten video yang ada didalamnya yang dapat mempermudah seseorang untuk mendapatkan keinginannya, dengan berbagai contoh konten yang ada didalam *youtube* adalah edukatif (pendidikan), *gaming*, vlog (*video blog*), musik, hiburan, berita (informasi), film, dan lain-lain. Dengan banyaknya konten yang terdapat didalam *youtube*, penonton dapat memilih sendiri tontonan yang ingin ditonton (Nurita, 2017), tidak seperti televisi pada umumnya, yang segala tayangannya sudah dijadwalkan sesuai dengan keinginan televisi.

Tidak bisa dipungkiri bahwa media sosial sudah “menjamur” dikalangan masyarakat ini. Mayoritas masyarakat sudah dapat mengakses internet dengan mudah, hal tersebut berdampak meledaknya atau banyaknya pengguna internet dari Indonesia. Berdasarkan survei di *We are Social*, menyatakan bahwa lebih dari 150 juta pengguna internet di Indonesia, 88% dari keseluruhannya yang berarti sekitar 132 juta orang merupakan pengguna aktif media sosial *Youtube*, selebihnya ke media sosial lain seperti *Whatsapp*, *Instagram*, *Line*, dan lain-

lain (katadata.co.id, 2019). Angka 88% dari keseluruhan pengguna aktif internet di Indonesia merupakan angka yang sangat amat besar, ditambah pengguna aktif *Youtube* di Indonesia ini merupakan media sosial dengan angka pengguna aktif tertinggi. Sedangkan dilansir dari press@google.com, dinyatakan bahwa *Youtube* memiliki lebih dari 1 miliar pengguna, hampir sepertiga dari dari semua pengguna internet didunia, dan disetiap harinya terjadi miliaran tayangan (Yolanda, 2018). Hal ini dapat menyimpulkan bahwa *Youtube* dapat memberikan peran yang

penting bagi trend di Indonesia maupun di dunia, dan lagi tingkat kepopuleran suatu *channel* yang ada didalamnya.

Dari data yang tersedia, kami akan mengelola data mengenai 5000 *channel youtube* yang kami ambil datanya pada bulan Januari 2019, agar dapat mengetahui popularitas atau grade yang didapatkan serta kriteria apa yang dibutuhkan untuk mendapatkan popularitas atau grade di atas "A" pada bulan Januari 2019. Berikut adalah sample atau beberapa *row data* tentang *channel youtube* yang ingin penulis gunakan :

Tabel 1. Sample Data (socialblade.com)

<i>Rank</i>	<i>Grade</i>	<i>Channel Name</i>	<i>Video Uploads</i>	<i>Subscribers</i>	<i>Video Views</i>
1 st	A++	Zee TV	82757	18752951	20869786591
2 nd	A++	T-Series	12661	61196302	47548839843
3 rd	A++	Cocomelon	373	19238251	9793305082
4 th	A++	SET India	27323	31180559	22675948293
5 th	A++	WWE	36756	32852346	26273668433

Berdasarkan dari Tabel 1 sample data yang diambil, terdapat variasi jenis data diantaranya adalah *Rank*, *Grade*, *Channel Name*, *Video uploads*, *Subscribers*, dan *Video views*. Namun *channel Youtube* yang berada di peringkat 1 pun memiliki jumlah *subscribers* atau pelanggan yang tidak lebih banyak dari *channel-channel* yang ada dibawahnya. Hal ini dapat menjadi pembahasan unik, bahwa jumlah *subscribers* bukan menjadi acuan utama dalam meraih popularitas di *Youtube*. Jika dilihat dari variabel lain yaitu *Video Uploads*, jumlah video yang diupload juga tidak dapat menjadi jaminan popularitas itu sendiri. Dari data sample yang diambil, dapat dilihat meski video yang diupload tidak terlalu banyak, namun jika konten yang ada didalamnya dapat menarik, membuat *channel* tersebut memiliki *grade* yang baik.

METODE PENELITIAN

Batasan Masalah dan Ruang Lingkup Penelitian ini dibatasi oleh beberapa hal diantaranya adalah :

Data yang diambil berupa dataset yang sudah berbentuk file excel (didownload langsung dari kaggle yang bersumber dari socialblade.com.

Penelitian ini bersifat mempelajari pemetaan grade didalam *channel youtube* berdasarkan dari dataset yang sudah dimiliki.

Ruang lingkup yang menjadi batas permasalahan ini mencakup seluruh *channel* di dunia, namun diambil dengan waktu

Mengambil Data, pengambilan data diambil dari *website kaggle*, yang didalamnya sudah terdapat data mengenai *channel youtube* beserta dengan grade nya.

Setelah melakukan pengambilan data, maka data ini perlu diimport atau di extract kedalam aplikasi pengelola data

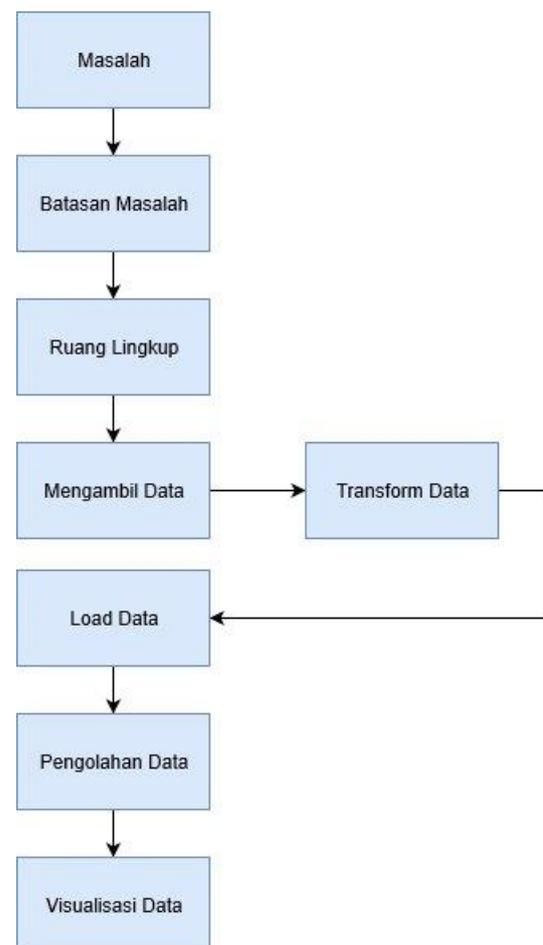
Transform Data, Dari *dataset* yang telah diambil atau didownload, didalamnya tentu terdapat variabel-variabel yang tidak memiliki nilai, maka dari itu *Transform Data* perlu untuk dilakukan. *Transform Data* berguna untuk membersihkan data agar sesuai dengan kriteria tidak terjadinya *error* ketika mengelola data,

Pengolahan Data, Didalam tahapan ini, data yang sudah ditransform atau dibersihkan dapat dilanjutkan ketahap pengolahan. Proses pengolahan data ini nantinya akan dijelaskan secara rinci dibagian pembahasan. Namun secara garis besar pengolahan data ini akan menggunakan aplikasi *Rapid Miner* sebagai alat untuk membantu proses pengolahan data.

Visualisasi Data, Tahap akhir dari pengolahan data ini adalah Visualisasi Data, hasil yang diharapkan nantinya yaitu berupa *visualisasi* ataupun diagram / *presentase* mengenai pemetaan channel youtube, agar pemetaan popularitas mengenai channel youtube dapat dipahami dengan mudah dan sederhana.

Di dalam pengelolaan data ini, kami ingin menggunakan algoritma ID3 sebagai metode pemetaan popularitas ini. Untuk pengertiannya, algoritma *Decision Tree* ataupun ID3 merupakan salah satu algoritma Klasifikasi yang merupakan Singkatan dari *Iterative Dichotomise 3* dan merupakan sebuah metode yang digunakan untuk membuat pohon keputusan yang telah dikembangkan oleh J. Ross Quinlan sejak tahun 1986. Algoritma pada metode ini menggunakan konsep dari entropy informasi. Konsep Dasar Metode ID3 Merupakan Metode Klasifikasi, dimana ID3 mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan *rule*, Data dalam ID3 biasanya dinyatakan dalam bentuk tabel yang memiliki atribut dan record. Atribut tersebut dapat digunakan sebagai Sebuah Parameter sebagai kriteria dalam Membuat Tree / Pohon Keputusan, Intinya konsep ID3 adalah mengubah data menjadi bentuk pohon (Arsitektur didalam ID3

menyerupai bentuk pohon), dan dari pohon ini akan diubah menjadi Rule dan pada prosesnya ID3 dimulai dari "root node" yang merupakan (Node akar) hingga proses akhirnya yaitu "leaf node" (Node daun). Dengan algoritma ini, kami akan meneliti popularitas dari suatu *grade* yang ada didalam *Youtube*.



Gambar 1. Metode Penelitian

HASIL DAN PEMBAHASAN

Di dalam pembahasan ini, tahap pertama yang kami lakukan adalah mengelompokkan terlebih dahulu channel berdasarkan masing-masing *grade* yang dimiliki dengan menggunakan aplikasi Power OLAP. Dengan menggunakan fungsi *Association*, maka kita dapat mengetahui berapa banyak *channel* yang memiliki suatu kategori *grade*, dengan tujuan dapat

mengetahui populasi *channel* didalam 1 kategori grade.

Kemudian difungsi berikutnya ada *Clustering*, didalam fungsi ini memproses grade berdasarkan alphabetnya yaitu grade A++, A+, A, A-. Tujuan kami menggunakan dua case ini adalah agar dapat mengetahui populasi *channel Youtube* yang memiliki suatu grade tertentu. Berikut adalah contoh *sample* dari *output* yang telah diproses oleh Power OLAP menggunakan metode *Association* diperlihatkan pada tabel 2 *Sample case association*.

Tabel 2. Sample Case Association

No.	Rank	Channel Name
1	A++	Zee TV
2	A++	T-Series
3	A++	Cocomelon – Nursery Rhymes
4	A++	SET India
5	A++	WWE
6	A++	Movieclips
7	A++	Netd mAzik
8	A++	ABS-CBN Entertainment
9	A++	Ryan ToysReview
10	A++	Zee Marathi

Ditahap berikutnya kami akan menggunakan aplikasi *Rapid Miner*, yang didalamnya kami akan menggunakan algoritma *Decision Tree* ID3. Didalam aplikasi *Rapid Miner* ini, kita perlu mengambil data mentah terlebih dahulu, bisa dalam bentuk file biasa ataupun lewat database, namun karena data yang kita miliki berformat *.xlsx / Excel*, maka perlu kita import via *File / My Computer*. Hal yang perlu kita lakukan untuk mengimport data tersebut adalah dengan mengklik tombol (+) *Import Data* didalam aplikasi *Rapid Miner*, berikut diperlihatkan pada Gambar 1. *Tampilan software Rapid miner*.

Setelah mengklik tombol tersebut, kita perlu mencari letak penyimpanan data mentah yang kita miliki. Kemudian setelah data mentah tersebut sudah dipilih, maka akan muncul tampilan dari data mentah yang kita pilih. Kita juga perlu menyesuaikan beberapa tipe variabel yang sudah ada, seperti contoh tipe-tipe variabel yang tersedia adalah *integer, numeric, text, time, date, polynominal, binominal*, dan lain-lain dengan tujuan agar pengolahan data didalam aplikasi ini dapat berjalan dengan baik dan hasil akhirnya dapat memberikan informasi yang sesuai. Setelah penyesuaian data sudah dilakukan, kita perlu menggunakan beberapa parameter untuk mengolah data mentah yang kita miliki, diantaranya beberapa parameter yang dibutuhkan adalah *Retrieve, Select Attribute*, dan ID3.



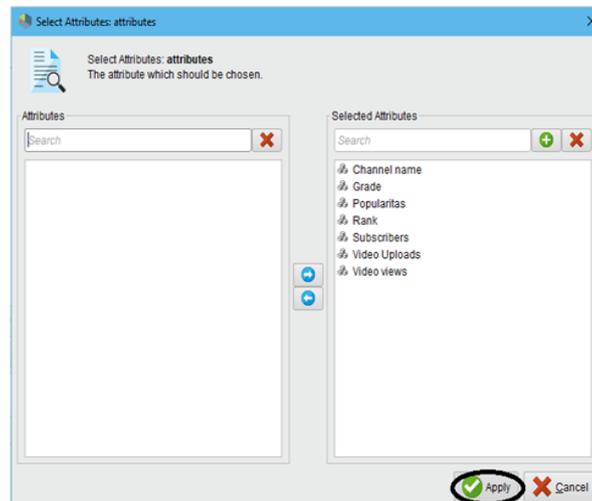
Gambar 2. Tampilan Software Rapid Miner.

Kegunaan *Retrieve* adalah untuk membaca data mentah yang kita miliki, kemudian disambungkan ke parameter *Select Attribute*. Kegunaan parameter *Select Attribute* ini adalah memilih variabel atau kolom yang kita butuhkan untuk diolah yang berasal dari data mentah yang kita miliki. Untuk memilih *variabel* atau kolom yang diinginkan, kita perlu mengubah "*attribute filter type*" didalam parameter *Select Attribute* ini menjadi *subset*, kemudian klik tombol "*Select Attributes...*", setelah itu

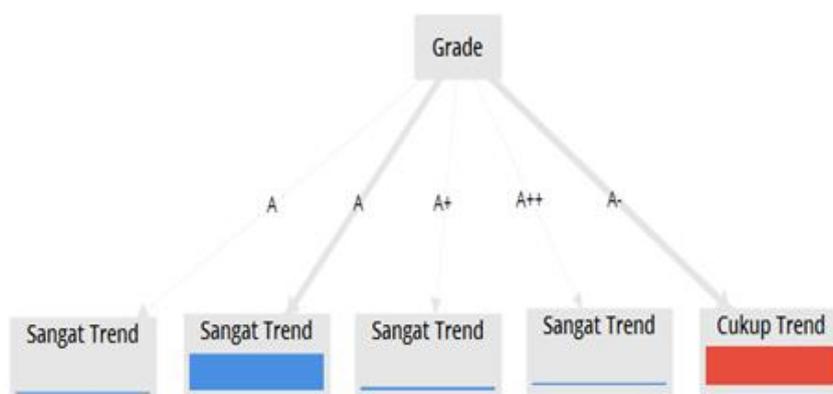
muncul lah pilihan variabel ataupun kolom dari data mentah yang kita miliki, kita hanya perlu menyesuaikan atau memilih variabel atau kolom mana yang ingin digunakan, jika sudah klik Apply, seperti diperlihatkan pada Gambar 3. Tampilan *Select Atribut*.

Jika sudah, maka parameter terakhir yang digunakan adalah ID3. Parameter ID3 ini yang nantinya akan menghasilkan output berupa tampilan pohon keputusan / decision tree mengenai data yang telah dipilih sebelumnya. Hal yang perlu kita lakukan setelah mengambil parameter ini adalah menyambungkan parameter sebelumnya yaitu *Select Attribute* ke parameter ID3 ini agar proses pengolahan data dapat dilakukan, jika

sudah maka sambungkan juga parameter ID3 ini ke *result*, agar hasil dari algoritma ID3 ini dapat menghasilkan pohon keputusan atau *Decision Tree*. Berikut adalah hasil dari pengolahan data menggunakan algoritma ID3, diperlihatkan pada Gambar 4. Hasil Pohon Keputusan.



Gambar 3. Tampilan Select Atribut



Gambar 4. Hasil Pohon Keputusan

Tree

Grade = A: Sangat Trend (Sangat Trend=1, Cukup Trend=0)
Grade = A : Sangat Trend (Sangat Trend=964, Cukup Trend=0)
Grade = A+ : Sangat Trend (Sangat Trend=40, Cukup Trend=0)
Grade = A++ : Sangat Trend (Sangat Trend=11, Cukup Trend=0)
Grade = A- : Cukup Trend (Sangat Trend=0, Cukup Trend=1024)

Gambar 5. Hasil Dari Pengelolaan Data

Pada Gambar 5 Hasil Dari Pengelolaan Data menggunakan algoritma ID3, ketebalan dari warna merupakan perwakilan dari banyaknya populasi yang ada didalam suatu grade, berdasarkan dari data yang dimiliki dan dari tree yang dihasilkan, dapat disimpulkan bahwa grade terbanyak merupakan A-, yaitu sebanyak 1024 channel, sedangkan populitas grade terbanyak dibawah A- adalah A, yaitu sebanyak 964 channel, dan dengan populitas popularitas yang paling sedikit yaitu A++ dan A+.

KESIMPULAN

Dari hasil pembahasan diatas dapat disimpulkan bahwa *subscribers*, *video uploads* dan *video views* dapat mempengaruhi popularitas dan grade dimana popularitas dapat didapat dari jumlah video yang ditonton, popularitas suatu akun *youtube* akan sangat dipengaruhi oleh seberapa banyak akun *youtube mengupload video* dimana semakin banyak video yang terupload akan semakin besar kemungkinan video ditonton, semakin banyaknya video yang tertonton dapat meningkatkan jumlah subscribers dari akun *youtube* itu sendiri dimana hal tersebut juga akan mempengaruhi perubahan pemetaan popularitas, semakin banyak orang yang berlangganan atau *subscribe* kepada suatu akun *youtube* tersebut maka akan semakin besar popularitas dari akun *youtube* tersebut, semakin

kecil maka semakin tidak populer akun *youtube* tersebut, 3 atribut ini (*subscribers*, *video uploads* dan *video views*) merupakan atribut yang saling terhubung dan mendukung peningkatan popularitas dari akun *youtube* itu sendiri.

DAFTAR PUSTAKA

- Cahyono, Anang Sugeng. (2016). "Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia."
- Dumbill, Edd. (2012). "Big Data Now: 2012 Edition. "What Is Big Data?" O'Reilly. USA: O'Reilly Media, Inc.
- Davis, Gordon B. (2012). "Computer Data Processing", McGraw-Hill Book Company; First Edition edition.
- Fatty, Faiqah, Muh. Nadjib. Andi Subhan Amir. (2016). "Youtube Sebagai Sarana Komunikasi Bagi Komunikasi Makassarvidgram."
- Nurita, Labas Yessi, Daisy Indira Yasmine. (2017). "Komodifikasi di Era Masyarakat Jejaring: Studi Kasus Youtube Indonesia."
- Muhaemin Ceceng, Iwan Krisnadi. (2015). "Pemanfaatan Big Data Melalui Media Sosial Untuk Pemetaan Keluhan Pelanggan".
- Yolanda, Stellarosa., Sandra Jasmine, Andre Ikhsano. (2018). "Pemanfaatan Youtube Sebagai Sarana Transformasi Majalah Highend."
- <https://socialblade.com/> (Sampel data youtube, diakses pada 23 Agustus 2019)
- <https://katadata.co.id/> (Survey media sosial, diakses pada 23 Agustus 2019).