

# Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5

<sup>1</sup>ARDI RAMADHAN SUKMA<sup>2</sup>RIQADRI HALFIS<sup>3</sup>ADY HERMAWAN

<sup>1</sup>Mahasiswa Pasca Sarjana, Ilmu Komputer , STMIK Nusa Mandiri, Indonesia

<sup>2</sup> Mahasiswa Pasca Sarjana, Ilmu Komputer , STMIK Nusa Mandiri, Indonesia

<sup>3</sup> Mahasiswa Pasca Sarjana, Ilmu Komputer , STMIK Nusa Mandiri, Indonesia

E-mail: [1ardirsukma40@gmail.com](mailto:1ardirsukma40@gmail.com) , [2arizhboys@gmail.com](mailto:2arizhboys@gmail.com) , [3ady.hermawan31@gmail.com](mailto:3ady.hermawan31@gmail.com)

## ABSTRAK

Pada zaman saat ini media sosial sangat berpengaruh dalam kehidupan kita entah dari Instagram, Twitter, Youtube, atau yang lainnya. Salah satunya Youtube media sosial yang punya peran sangat penting bagi kalangan masyarakat walaupun public figure. Youtube menyediakan content yang sangat menarik bagi yang mempunyai channel karena memiliki penilai dari Video Upload, Subscriber, Video Viewers. Bahkan Youtube memberikan peluang bagi channel Youtube yang memasang Google AdSense untuk dibayarkan kepada pemilik Channel tersebut. Penilaian Channel Youtube bisa di nilai dari Video Upload, Subscriber, Video Viewers. Dari data Youtube tersebut dapat dilakukan pengolahan data. Salah satu teknik pengolahan data yang dapat digunakan dalam proses tersebut adalah klasifikasi. Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Dan menggunakan Algoritma C4.5 dalam proses pengklasifikasi-annya. Yang dapat menentukan Channel Youtube khususnya Indonesia dengan ratio Sangat Baik atau Baik.

**Keywords:** *Algoritma C4.5, Indonesia, Youtube, Data Mining, Klasifikasi*

## 1. Pendahuluan

### 1.1 Latar Belakang

Pada zaman saat ini perkembangan media sosial khususnya di Indonesia sangat pesat, berbagai informasi dapat di terima dengan media sosial. Media sosial merupakan sebuah media daring, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Banyak sekali jejaring sosial di dunia maya salah satu nya Youtube.

Youtube merupakan sebuah situs web berbagi video yang dibuat oleh tiga mantan karyawan PayPal pada Februari 2005. Situs ini memungkinkan pengguna mengunggah, menonton, dan berbagi video. Perusahaan ini berkantor pusat di San Bruno, California, dan memakai teknologi Adobe Flash

Video dan HTML5 untuk menampilkan berbagai macam konten video buatan pengguna, termasuk klip film, klip TV, dan video musik. Selain itu ada pula konten amatir seperti blog video, video orisinal pendek, dan video pendidikan. Di Youtube bisa mendapatkan penghasilan dari Google AdSense atau pun Video Upload, Subscriber, Video Viewers. Dengan Metode C4.5 saya akan penelitian dengan atribut Category, Video Upload, Subscriber, Video Viewers, dan Ratio.

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross quinlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja, yaitu: Membuat pohon keputusan, dan membuat aturan-aturan

(*rule model*). Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

## 1.2 Rumusan Masalah

Bagaimana menerapkan Algoritma untuk mengklasifikasikan Channel Youtube Indonesia ?

## 1.3 Tujuan Penelitian

Untuk mengetahui klasifikasi channel youtube Indonesia dengan melihat ratio menggunakan Algoritma C4.5.

## 1.4 Metode Penelitian

- a. Melakukan pengumpulan data Channel Youtube Indonesia.
- b. Menggunakan Algoritma C4.5 untuk klasifikasi Channel Youtube Indonesia.

## 2. Landasan Teori

### 2.1 Data Mining

Data mining yang di kenal dengan nama *pattern recognition* adalah metode pengolahan untuk menemukan suatu pola yang tersembunyi untuk dapat di olah menjadi pengetahuan dan ilmu pengetahuan baru dan informasi dari data dan hasil untuk keputusan di masa depan.

Data mining juga dapat di sebut sebagai sistem pengolahan data yang sangat besar, yang memberikan peranan dalam beberapa bidang di dunia yaitu bidang keuangan, industri, transportasi, cuaca, dan teknologi. Dalam data mining juga terdapat metode metode yang dapat di gunakan antara lain metode klasifikasi, *clustering*, *regresi*, seleksi *variabel*, dan *market bisnis*. Data mining dapat di artikan sebagai data dalam jumlah besar yang di simpan dalam suatu database, *data warehouse* untuk menyimpan dapat sehingga dapat menemukan pola. Ada beberapa teknik dalam data mining antara lain data analisis, *signal processing*, *neural network* dan pengenalan pola.

### 2.2 Fungsi Data Mining

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu:

1. **Fungsi Prediksi (prediction).** Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.
2. **Fungsi Deskripsi (description).** Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.
3. **Fungsi Klasifikasi (classification).** Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.
4. **Fungsi Asosiasi (association).** Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

### 2.3 Tahapan Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap preprocessing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut (Fayyad, 1996):

#### 1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data

hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

## 2. Pre-processing / cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

## 3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

## 4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

## 5. Interpretation / evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 2.4 Proses Data Mining

Proses yang umumnya dilakukan oleh data mining antara lain: deskripsi, prediksi,

estimasi, klasifikasi, clustering dan asosiasi. Secara rinci proses data mining dijelaskan sebagai berikut (Larose, 2005):

#### *a. Deskripsi*

Deskripsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (knowledge) pada sistem. Tugas deskriptif merupakan tugas data mining yang sering dibutuhkan pada teknik postprocessing untuk melakukan validasi dan menjelaskan hasil dari proses data mining. Postprocessing merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

#### *b. Prediksi*

Prediksi memiliki kemiripan dengan klasifikasi, akan tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang.

#### *c. Estimasi*

Estimasi hampir sama dengan prediksi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi.

#### **d. Klasifikasi**

Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya.

#### **e. Clustering**

Clustering merupakan pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama. Sebuah kluster adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu cluster dan semakin besar perbedaan tiap cluster maka kualitas analisis cluster semakin baik.

#### **f. Asosiasi**

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analysis). Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

### 2.5 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross Quinlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja, yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*). Aturan aturan yang terbentuk

dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

1. Memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
2. Membuat cabang untuk masing-masing nilai, artinya membuat cabang sesuai dengan jumlah nilai variabel gain tertinggi.
3. Membagi setiap kasus dalam cabang, berdasarkan perhitungan nilai gain tertinggi dan perhitungan dilakukan setelah perhitungan nilai gain tertinggi awal dan kemudian dilakukan proses perhitungan gain tertinggi kembali tanpa meyeritakan nilai variabel gain awal.
4. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama, mengulangi semua proses perhitungan gain tertinggi untuk masing-masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Dari ketiga peneliti yang telah dilakukan tersebut, klasifikasi dengan Algoritma C4.5 digunakan oleh para peneliti sebagai solusi untuk mengambil keputusan yang diharapkan mampu membantu dalam pengambilan keputusan dengan lebih mudah dan cepat.

Dalam penerapan dan penggunaan algoritma C4.5, dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon pegawai yang berpotensi untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan. Selain itu algoritma ini digunakan untuk klasifikasi predikat

keberhasilan mahasiswa disebut universitas. Variabel yang memiliki prioritas utama terhadap predikat keberhasilan mahasiswa adalah mahasiswa yang memilih sesi perkuliahan pada pagi hari.

## 2.6 Entropy Dan Information Gain

*Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity dan homogeneity dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* masing-masing atribut. Penghitungan nilai *Entropy* digunakan rumus seperti dalam Persamaan (1).

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i) \dots\dots\dots(1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan entropy yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya :

- s : Himpunan kasus
- n : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam s

*Information Gain* adalah informasi yang didapatkan dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu set data.

$$GAIN(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \dots\dots(2)$$

## 2.8 Landasan Penelitian

### a. Jurnal Internasional

Parameter Riset	Perbedaan Riset		
	Kalpesh Adhartao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao (2013)	I.A Ganiyu (2016)	Seema, Jitendra Agrawal, Sanjeev Sharma (2013)
Pemakaian Metode/Algoritma	ID3 dan C4.5	C4.5	C4.5
Objek Penelitian	Students Performance	Academic Performance	Various Entropies
Output	Memprediksi Performance Student	Prediksi Performance Pelajar Science menggunakan Kalsifikasi	Klasifikasi Berbagia Entropi

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *information gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya :

- s : Himpunan kasus
- n : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam s

Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*.

## 2.7 Pohon Keputusan (*Decision Tree*)

Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk mengubah data menjadi pohon keputusan yang akan menghasilkan aturan-aturan keputusan besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pohon keputusan yang dihasilkan oleh algoritma C4.5 dapat digunakan untuk klasifikasi. Pohon keputusan adalah salah satu teknik penambangan data yang paling populer untuk penemuan pengetahuan. Secara sistematis menganalisis dan mengekstrak aturan untuk tujuan Klasifikasi / prediksi.

Judul	Predicting Students' Performance Using ID3 And C4.5 Classification Algorithms	Data Mining : A Prediction for Academic Performance Improvement of Science Students using Classification	Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies
-------	---	--	--

b. Jurnal Nasional

Parameter Riset	Perbedaan Riset	
	Mardi Turnip, Charles Wijaya (2016)	Aradea, Satriyo A., Ariyan Z., Yaliana A. (2011)
Pemakaian Metode/Algoritma	C4.5	C4.5
Objek Penelitian	Tingkat Konsumsi Konsumen	Penentuan Penerimaan Beasiswa
Output	Menyediakan informasi yang diperlukan pihak manajemen, klasifikasi tingkat konsumsi konsumen	Klasifikasi Penentuan Pola PMB
Judul	Penerapan Algoritma C4.5 untuk Penentuan Tingkat Konsumsi Konsumen pada Medan Solusindo	Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru

2.9 Objek Penelitian

Objek Penelitian adalah Channel Youtube Indonesia berdasarkan Video Upload, Subscribers, Video Viewers, dan Ratio dimana Ratio sebagai label untuk mengklasifikasi.

3. Metode Penelitian

Metode Penelitian yang digunakan dalam penerapan algoritma C4.5 untuk klasifikasi Channel Youtube Indonesia, menggunakan rancangan penelitian sebagai berikut :

1. Pengumpulan Data  
Pengumpulan data adalah mengumpulkan data-data yang akan digunakan dalam proses algoritma klasifikasi C4.5.
2. Transformasi dan Seleksi Data  
Transformasi data adalah proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai, agar dapat di proses dengan perhitungan algoritma C4.5. Transformasi Data adalah proses mentransformasi atau menggabungkan data ke dalam bentuk yang sesuai untuk

penggalan lewat operasi summary atau aggregation.

3. Perhitungan *Entropy* dan *Information Gain*  
Perhitungan semua atribut/variabel, *entropy* menggunakan rumus (1) dan *information gain* menggunakan rumus (2) untuk mengetahui *information gain* tertinggi yang akan dijadikan simpul akar pada pembuatan pohon keputusan.
4. Pohon Keputusan (Decision Tree)  
Pohon keputusan adalah hasil dari proses perhitungan *entropy* dan *information gain*, setelah perhitungan berulang-ulang sampai semua atribut pohon memiliki kelas dan tidak bisa lagi dilakukan proses perhitungan.
5. Rule Model  
Aturan-aturan/ Rule model adalah uraian penjelasan yang merepresentasikan sebuah pohon keputusan.
6. Validasi dan Pengujian  
Validasi dan pengujian adalah Pengujian yang dilakukan untuk mengetahui semua

fungsi bekerja dengan baik atau tidak. Validasi dilakukan dengan *Ten-fold Cross Validation*. *Ten-fold Cross Validation* adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah persentase dari catatan yang diklasifikasikan dengan benar dalam pengujian dataset. Presisi adalah persentase data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. Recall adalah pengukuran tingkat pengenalan positif sebenarnya.

## 4. Hasil dan Pembahasan

### 4.1 Pengumpulan Data

Pengumpulan Data dilakukan dengan cara mengambil di website [socialblade.com/youtube/top/country/id](https://socialblade.com/youtube/top/country/id). Data ini mencakup beberapa variable, antara lain : Rank, Grade, Channel name, Video Upload, Subscriber, Video view, yang akan diseleksi dengan algoritma C4.5. Jumlah data digunakan sebanyak 110 data Channel Youtube Indonesia.

Tabel 1. Data Channel Youtube Indonesia yang akan dijadikan data, dalam proses algoritma klasifikasi C4.5

Rank	Grade	Channel name	Video Upload	Subscriber	Video view
1 <sup>st</sup>	A	Indosiar	18994	4111015	2533563033
2 <sup>nd</sup>	A	RCTI - Layar Drama Indonesia	29466	3334422	2879041290
3 <sup>rd</sup>	A	TRANSTV Official	27115	3351332	1909556996
...	...	...	...	...	...
110 <sup>th</sup>	B+	Sony Entertainment Indonesia	836	610916	679908981

### 4.2 Transformasi Data

Transformasi data dilakukan dengan mengubah beberapa nilai atribut yang awalnya bernilai nominal menjadi nilai-nilai atribut

yang sesuai dengan data pada Tabel 2 agar dapat dilakukan proses perhitungan algoritma klasifikasi C4.5.

Tabel 2. Tabel Transformasi data

Channel name	Category	Video Upload	Subscriber	Video View	Ratio
Indosiar	TV	Menengah	Tinggi	Tinggi	Sangat Baik
RCTI - Layar Drama Indonesia	TV	Tinggi	Tinggi	Tinggi	Sangat Baik
TRANSTV Official	TV	Tinggi	Tinggi	Tinggi	Sangat Baik
...	...	...	...	...	...
Sony Entertainment Indonesia	Entertainment	Rendah	Rendah	Menengah	Baik

#### 4.3 Perhitungan *Entropy* dan *Information Gain*

Dengan cara menggunakan rumus *Entropy* kita dapat melakukan perhitungan jumlah keputusan “Sangat Baik” dan “Baik” dari seluruh kasus yang ada.

$$Entropy(Total) = (-24/110) \cdot \log_2(24/110) + (-86/110) \cdot \log_2(86/110) = 0.7569834$$

Jika sudah mendapatkan nilai *Entropy* dari tiap-tiap variable yang ada seperti Category, Video Upload, Subscriber, Video Viewers, Maka langkah selanjutnya menghitung *Information Gain* pada setiap atributnya.

$$Gain(Total, Category) = (0.7569834) - ((4/110) \cdot 0) - ((30/110) \cdot 0.650022) - ((23/110) \cdot 0.558629) - ((32/110) \cdot 0.696212) - ((16/110) \cdot 0.954434) = 0.121389$$

$$Gain(Total, Video Upload) = (0.7569834) - ((9/110) \cdot 0.503258) - ((4/110) \cdot 0.811278) - ((97/110) \cdot 0.568363) = 0.184965$$

$$Gain(Total, Subscriber) = (0.7569834) - ((14/110) \cdot 0.591673) - ((42/110) \cdot 0.775135) - ((54/110) \cdot 0.309543) = 0.233611$$

$$Gain(Total, Video Viewers) = (0.7569834) - ((17/110) \cdot 0.522559) - ((77/110) \cdot 0.520335) - ((16/110) \cdot 0) = 0.311184$$

Setelah semua *Entropy* dan *Information Gain* di hitung maka hasil perhitungan di masukan dalam tabel 3. Dari perhitungan di atas di ketahui bahwa *Information gain* terbesar adalah Video Viewers, maka dapat di jadikan simpul akar. Lalu ambil *Information Gain* terbesar yaitu Video Viewers lakukan perhitungan ulang dengan cara menghilangkan atribut yang terbesar tadi hingga semua atribut pohon memiliki kelas.

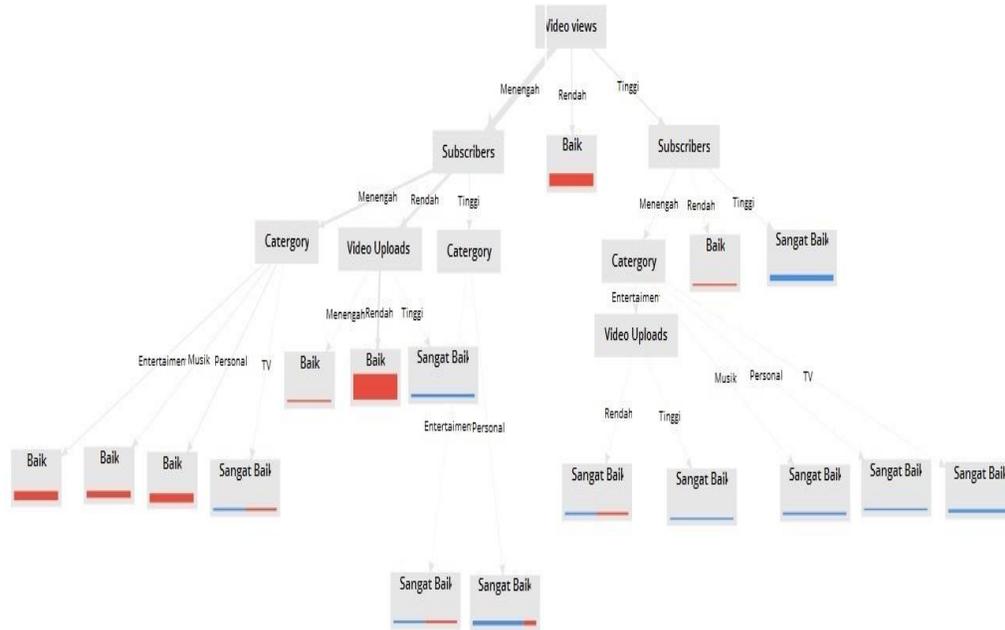
Tabel 3. Hasil Perhitungan *Entropy* dan *Information Gain*

		jumlah	Sangat Baik	Baik	Entropy	Gain
total		110	24	86	0.756834	
Category						0.121389
	E-commerce	4	0	4	0	
	Entertainment	30	5	25	0.650022	
	Musik	23	3	20	0.558629	
	Personal	32	6	26	0.696212	
	TV	16	10	6	0.954434	
Video Uploads						0.184965
	Tinggi	9	8	1	0.503258	
	Menengah	4	3	1	0.811278	
	Rendah	97	13	84	0.568363	
Subscriber						0.233611
	Tinggi	14	12	2	0.591673	
	Menengah	42	9	32	0.775135	
	Rendah	54	3	51	0.309543	
Video Views						0.311184
	Tinggi	17	15	2	0.522559	
	Menengah	77	9	68	0.520335	
	Rendah	16	0	16	0	

#### 4.4 Pohon Keputusan (Decision Tree)

Dari hasil Perhitungan *entropy* dan *information gain* yang didapat kemudian

diolah kedalam *Decision Tree*. Berikut hasil Pohon Keputusan (*Decision Tree*) :



Gambar 1. Pohon Keputusan (*Decision Tree*)

#### 4.5 Rule Model

Dari pohon keputusan (*decision tree*) yang terbentuk pada Gambar 1 diatas, didapat aturan—aturan/*rule*. Ada beberapa aturan yang terbentuk, dapat dilihat sebagai berikut ;

1. *If* Video Viewers=Rendah *then* Ratio = Baik
2. *If* Video Viewers=Menengah, and Subscriber=Menengah, Category=Entertainment *then* Ratio = Baik
3. *If* Video Viewers=Menengah, and Subscriber=Menengah, Category=Musik *then* Ratio = Baik
4. *If* Video Viewers=Menengah, and Subscriber=Menengah, Category=Personal *then* Ratio = Baik
5. *If* Video Viewers=Menengah, and Subscriber=Menengah, Category=TV *then* Ratio = Sangat Baik
6. *If* Video Viewers=Menengah, and Subscriber=Rendah, Video Uploads=Menengah *then* Ratio=Baik
7. *If* Video Viewers=Menengah, and Subscriber=Rendah, Video Uploads=Rendah *then* Ratio=Baik
8. *If* Video Viewers=Menengah, and Subscriber=Rendah, Video Uploads=Tinggi *then* Ratio=Sangat Baik
9. *If* Video Viewers=Menengah, and Subscriber=Tinggi, Category=Entertainment *then* Ratio=Sangat Baik
10. *If* Video Viewers=Menengah, and Subscriber=Tinggi, Category=Personal *then* Ratio=Sangat Baik
11. *If* Video Viewers=Tinggi, and Subscriber=Tinggi *then* Ratio=Sangat Baik
12. *If* Video Viewers=Tinggi, and Subscriber=Rendah *then* Ratio=Baik

13. *If* Video Viewers=Tinggi, *and* Subscriber=Menengah, Category=Entertainment, Video Uploads=Rendah *then* Ratio=Sangat Baik
14. *If* Video Viewers=Tinggi, *and* Subscriber=Menengah, Category=Entertainment, Video Uploads=Tinggi *then* Ratio=Sangat Baik
15. *If* Video Viewers=Tinggi, *and* Subscriber=Menengah, Category=Musik *then* Ratio=Sangat Baik

- Category=Personal *then* Ratio=Sangat Baik
17. *If* Video Viewers=Tinggi, *and* Subscriber=Menengah, Category=TV *then* Ratio=Sangat Baik

#### 4.6 Validasi dan Pengujian

Pengujian dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah *cross validation*. Berikut hasil dari *cross validation* :

16. *If* Video Viewers=Tinggi, *and* Subscriber=Menengah,  
accuracy: 92.73% +/- 6.80% (micro average: 92.73%)

	true Sangat Baik	true Baik	class precision
pred. Sangat Baik	21	5	80.77%
pred. Baik	3	81	96.43%
class recall	87.50%	94.19%	

Gambar 2. Hasil *Cross Validation*

Berdasarkan hasil pengujian dengan menggunakan *Cross-Validation* maka di dapatkan nilai akurasi 92.73% dengan class precesion Sangat Baik 80.77% dan class precision Baik 96.43%, serta class recall Sangat Baik 87.50% dan class recall Baik 94.19%.

#### 5. Kesimpulan

Kesimpulan dari penelitian ini adalah :

1. Penerapan Algoritma C4.5 akan dapat diimplementasikan pada ratio Channel Youtube Indonesia , dilihat dari akurasi yang mencapai 92.73%, bahwa perhitungan yang dilakukan dapat memprediksi dan meklasifikasi ratio pada Youtube Channel Indonesia.
2. Terbentuk beberapa aturan atau rule model yang dapat digunakan sebagai acuan dalam membuat aplikasi GUI.

#### REFERENCES:

- [1] Kalpesh Adhartao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao, "Predicting Students' Performance Using ID3 And C4.5 Classification Algorithms", *International Journal of Data Mining & Knowledge Management Process*, Departement of Computer Engineering(India), Vol.3, No.5, 2013, pp. 39-52.
- [2] I.A Ganiyu, "Data Mining : A Prediction for Academic Performance Improvement of Science Students using Classification", *International Journal of Information and Communication Technology Research*, Vol. 6, No. 4, 2016, ISSN 2223-4985.
- [3] Seema, Jitendra Agrawal, Sanjeev Sharma, "Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies", *International Journal of Computer Application*, Vol. 82, No. 16, 2013, pp 20-27.

- [4] Mardi Turnip, Charles Wijaya, "Penerapan Algoritma C4.5 untuk Penentuan Tingkat Konsumsi Konsumen pada Medan Solusindo", *Jurnal Senopati (Seminar Nasional Pascasarjana Teknik Informatika)*, Vol. 1, No. 1, 2016, pp. 34-42, p-ISSN 2528-2832, e-ISSN 2549-7820.
- [5] Aradea, Satriyo A., Ariyan Z., Yaliana A., "Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru", *Jurnal Penelitian Sitotika*, Vol. 7, No. 1, 2011, ISSN 1693-9670.
- [6] SOCIALBLADE, "Track User Statistic", website <https://socialblade.com/>.