

Penerapan Pohon Keputusan Pada Analisis Sentimen Media Sosial Twitter

Studi Kasus: Trailer Film *Avengers: Infinity War*

Christian Wibisono - 13516147
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13516147@std.stei.itb.ac.id

Abstract—Seiring berkembangnya media jejaring sosial seperti Twitter, semakin banyak hal yang dapat dilakukan untuk mendayagunakan data berbasis teks yang dihasilkan oleh seluruh masyarakat di dunia. Salah satu pemanfaatan jejaring sosial ini adalah untuk menganalisis dan memetakan kecenderungan sikap pengguna media sosial terhadap suatu topik tertentu. Analisis sentimen ini dapat dilakukan dengan menerapkan konsep pohon keputusan. Dalam makalah ini, penulis akan menganalisis sentimen masyarakat pada media sosial Twitter mengenai topik trailer film *Avengers: Infinity War* dengan menggunakan metode pohon keputusan dan pemanfaatannya untuk perumusan strategi pengembangan produk.

Kata Kunci—analisis sentimen, avengers, media sosial, pohon keputusan

I. PENDAHULUAN

Analisis sentimen atau *opinion mining* adalah proses memahami, mengurai, dan mengolah data tekstual secara komputasional untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen digunakan untuk memetakan kecondongan opini dan sikap masyarakat terhadap suatu masalah atau topik. Analisis sentimen ini mengelompokkan opini menjadi tiga kelompok besar yaitu positif, negatif, dan netral. Di era industri saat ini, penerapan analisis sentiment telah membawa pengaruh besar dalam kemajuan dan peningkatan kualitas layanan industri. Oleh sebab itu, tidak heran apabila penelitian di bidang ini semakin marak dan berkembang pesat.

Twitter merupakan salah satu media sosial yang mendukung dan menyediakan *open data API* yang bisa diakses oleh pengembang dan peneliti untuk melakukan *text mining*. Pengguna aktif Twitter pada 2017 ini tercatat masih mencapai sekitar 328.000.000,00 di seluruh dunia⁵. Oleh sebab itu data yang diperoleh dari Twitter dinilai cukup relevan dan bisa menggambarkan kondisi masyarakat yang sebenarnya. Data yang diperoleh dari Twitter dapat dikategorikan menurut *hashtag*('#') atau tanda pagar yang merujuk ke suatu topik tertentu.

Pada 30 November 2017, publik digemparkan dengan rilisnya trailer perdana film *Avengers: Infinity War*. Film yang dibuat Marvel Studios ini bahkan memecahkan rekor dengan menjadi trailer film yang paling banyak disaksikan di Youtube dengan

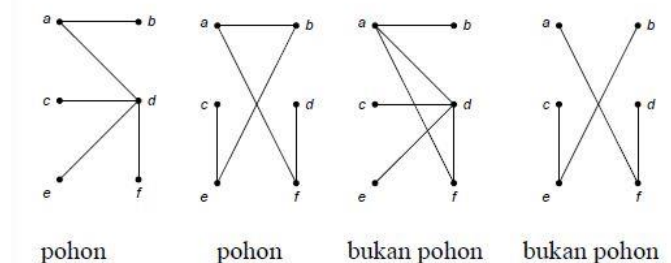
230.000.000,00 penonton dalam 24 jam. Media sosial Twitter turut diramaikan dengan berbagai komentar, opini, dan respon masyarakat mengenai trailer film ini. Topik mengenai film ini langsung menjadi *trending topic* dunia hanya dalam waktu singkat sejak trailernya dirilis ke publik. Trailer *Infinity War* juga menjadi trailer pertama film yang paling banyak dibahas pada Twitter dengan 1.147.438 *mention* dalam 7 jam pertama setelah dirilis⁴.

Dalam makalah ini penulis akan melakukan analisis sentimen media sosial twitter terhadap trailer film *Avengers: Infinity War* dengan mengimplementasikan program menggunakan pohon keputusan. Analisis ini kemudian dapat digunakan sebagai tolok ukur keberhasilan suatu film dalam melakukan pemasaran dan media pembelajaran strategi pemasaran melalui media digital.

II. LANDASAN TEORI

A. Pohon

Struktur pohon merupakan struktur yang penting dalam bidang matematika diskrit. Struktur pohon memungkinkan kita untuk mengorganisasi dan mengelola informasi berdasarkan suatu logika tertentu. Definisi Pohon dilihat sebagai struktur diskrit adalah graf tak berarah-terhubung yang tidak mengandung sirkuit.^[2] Sesuai dengan definisi tersebut ada dua sifat penting yang harus dimiliki pohon, terhubung dan tidak mengandung sirkuit



Gambar 2.1 Graf

Sumber: [http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2017-2018/Pohon%20\(2013\).pdf](http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2017-2018/Pohon%20(2013).pdf) diakses pada 30 November 2017 pukul 09.40 WIB

Pada gambar di atas G3 dan G4 bukan termasuk pohon. G3 mengandung sirkuit sedangkan G4 tidak terhubung. Pohon

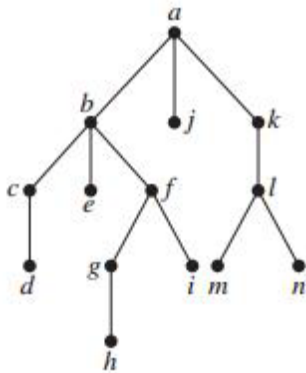
memiliki sifat-sifat yang dinyatakan di bawah ini^[2]. Misalkan graf $G = (V,E)$ adalah graf tak-berarah sederhana dan jumlah simpulnya n . Maka semua pernyataan di bawah ini ekuivalen :

1. G adalah pohon.
2. Setiap pasang simpul di dalam G terhubung dengan lintasan tunggal.
3. G terhubung dan memiliki
4. G tidak mengandung sirkuit dan memiliki sisi sebanyak $m = (n-1)$ buah.
5. G tidak mengandung sirkuit dan penambahan satu sisi pada graf akan menyebabkan munculnya satu sirkuit.
6. G terhubung dan semua sisinya adalah jembatan. (jembatan adalah sisi yang bila dihapus menyebabkan graf terpecah menjadi dua komponenn).

Ada berbagai macam jenis pohon, namun yang akan dipergunakan dalam implementasi pohon keputusan adalah pohon berakar.

B. Pohon Berakar

Pohon yang suatu simpulnya diperlakukan sebagai akar dan sisi-sisinya diberi arah sehingga menjauh dari akar dan menjadi graf berarah dinamakan pohon berakar (*rooted tree*).



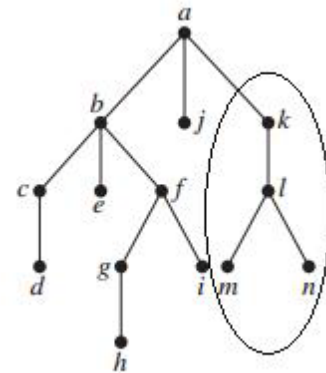
Gambar 2.2 Pohon Berakar
 Sumber: *Discrete Mathematics and Its Applications, Kenneth H. Rosen, Page 753*

Pohon berakar memiliki beberapa terminologi penting antara lain²:

1. Akar
 Akar adalah sebuah simpul yang memiliki derajat-masuk sama dengan nol. Artinya akar merupakan permulaan dari sebuah pohon dan hanya mempunyai sisi yang keluar dari simpul.
2. Anak (*child*) dan Orangtua (*parent*)
 Jika ada simpul x dan y , maka y dikatakan anak dari x jika ada sisi dari x ke y . Begitupula x disebut dengan orangtua dari y .
3. Lintasan (*path*)
 Lintasan dari simpul x_1 ke x_y adalah runtutan simpul-simpul dari $x_1, x_2, x_3, \dots, x_y$ sedemikian sehingga x_i

adalah orangtua dari x_{i+1} dengan $1 \leq i \leq y$. Panjang lintasan adalah jumlah sisi yang dilalui, yaitu $y - 1$.

4. Keturunan (*descendant*) dan Leluhur (*ancestor*)
 Jika terdapat lintasan dari simpul x ke y maka x merupakan leluhur dari y dan y merupakan keturunan dari x .
5. Saudara kandung (*sibling*)
 Dua buah simpul atau lebih dikatakan sebagai saudara kandung jika simpul-simpul tersebut memiliki orangtua yang sama.
6. Upapohon (*subtree*)



Gambar 2.3 Upapohon
 Sumber: *Discrete Mathematics and Its Applications, Kenneth H. Rosen, Page 753*

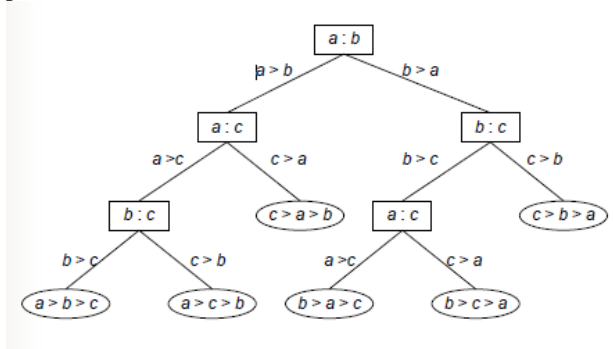
Jika x merupakan sebuah simpul di dalam pohon T , maka upapohon dengan menjadikan simpul x sebagai akarnya merupakan sebuah upagraf $T' = (V', E')$ sedemikian sehingga V' mengandung x dan seluruh keturunannya dan E' mengandung seluruh lintasan yang berasal dari x . Dalam satu pohon terdapat banyak sekali upapohon yang dapat dibentuk.

7. Derajat (*degree*)
 Derajat pada pohon punya definisi yang berbeda dengan graf. Derajat sebuah simpul pada pohon berakar adalah jumlah upapohon (atau jumlah anak) yang dimiliki oleh simpul tersebut.
8. Daun (*leaf*)
 Simpul yang mempunyai derajat keluar nol (tidak mempunyai anak) disebut dengan daun.
9. Simpul dalam (*internal nodes*)
 Simpul pada pohon yang mempunyai anak dan bukan akar merupakan simpul dalam.
10. Aras (*level*) atau Tingkat
 Aras dari akar adalah 0. Untuk simpul selain akar maka arasnya adalah $1 +$ panjang lintasan dari akar ke simpul tersebut.

11. Tinggi (*height*) atau Kedalaman (*depth*)
 Aras maksimum dari suatu pohon disebut sebagai tinggi.

C. Pohon Keputusan

Salah satu penerapan pohon berakar adalah pohon keputusan. Pohon keputusan adalah suatu pemodelan dari sebuah persoalan yang terdiri dari rangkaian keputusan yang mengarah kepada solusi dari persoalan tersebut. Setiap simpul menyatakan sebuah keputusan yang perlu dipertimbangkan, sedangkan daun menyatakan solusi dari permasalahan.



Gambar 2.4 Pohon Keputusan Untuk Mengurutkan 3 buah elemen

Sumber: Matematika Diskrit, Rinaldi Munir, Hal. 479

D. Analisis Sentimen

Analisis Sentimen / Opinion Mining merupakan sebuah cabang penelitian di domain Text Mining yang sedang mengalami perkembangan yang cukup pesat. Riset dan penelitian mengenai hal ini dimulai oleh Bo Pang (2002). Pada papernya, Bo Pang melakukan klasifikasi sentimen terhadap review film dengan menggunakan berbagai teknik pembelajaran mesin. Secara umum, analisis sentiment terbagi menjadi dua kategori besar yaitu:

1. *Coarse-grained sentiment analysis* - kita mencoba melakukan proses analisis pada level Dokumen. Singkatnya adalah kita mencoba mengklasifikasikan orientasi **sebuah dokumen** secara keseluruhan. Orientasi ini ada 3 jenis : **Positif, Netral, Negatif**. Akan tetapi, ada juga yang menjadikan nilai orientasi ini bersifat kontinu / tidak diskrit.
2. *Fined-grained sentiment analysis* - Obyek yang ingin diklasifikasi bukan berada pada level dokumen melainkan sebuah **kalimat** pada suatu dokumen.
 contoh :
 - Saya kecewa dengan pelayanan yang diberikan hotel tersebut. (**negatif**)
 - Saya sangat suka Matematika Diskrit. (**positif**)
 - Menurut saya, apa yang dia lakukan biasa saja (**netral**)

III. PENERAPAN POHON KEPUTUSAN PADA ANALISIS SENTIMEN

A. Data Preprocessing

Proses yang pertama dilakukan dalam analisis sentimen menggunakan pohon keputusan adalah menentukan membersihkan sebuah data. Dalam kasus ini data yang kita peroleh adalah data teks yang bersifat *unstructured*. Oleh sebab itu data ini harus dibersihkan dengan menghilangkan simbol-simbol yang tidak relevan, pranala luar untuk dianalisis menggunakan regex statements.

```
def clean_tweet(self, tweet):
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([\^0-9A-Za-z \t])|(\w+:\//\//S+)", "", tweet).split())
```

Sampel Data:

```
OH MYYY GOOOOODDDDD IM SO EXCITED FOR THIS PLEASE :))))
NME: Did you notice these characters were missing from the new #Avengers trailer?
https://t.co/ZLTryhW0Qe
```

Hasil Cleansing:

```
OH MYYY GOOOOODDDDD IM SO EXCITED FOR THIS PLEASE
NME: Did you notice these characters were missing from the new #Avengers trailer?
```

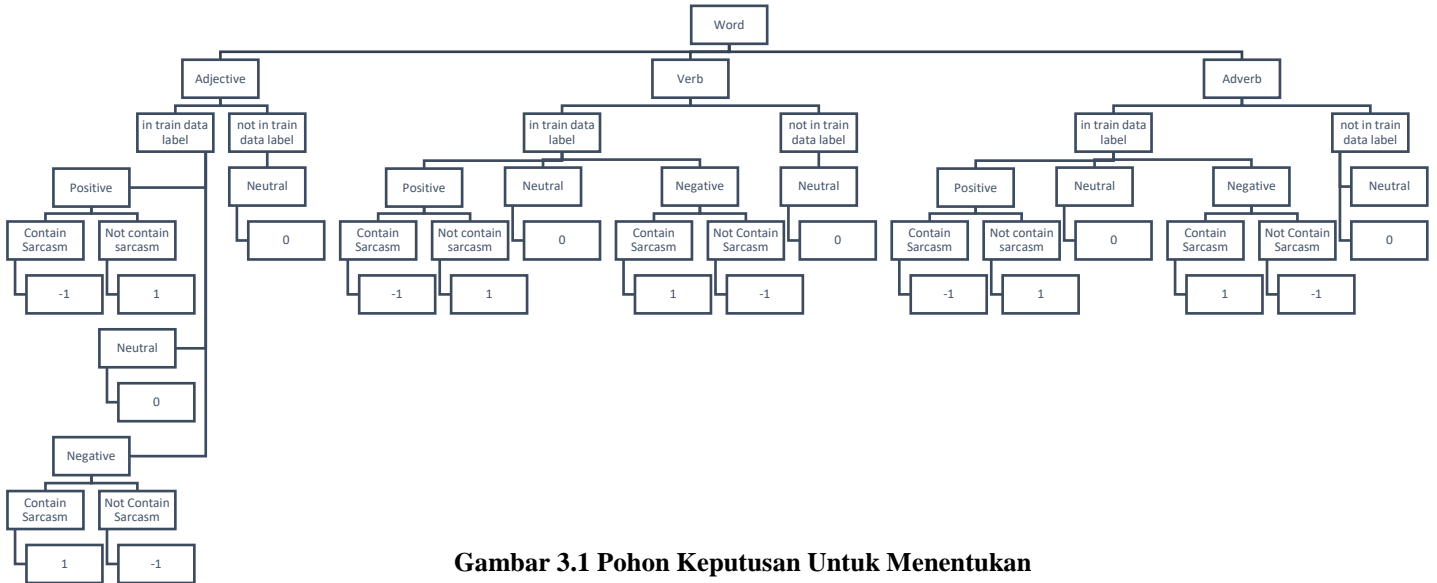
B. Data Labelling

Setelah data teks *unstructured* ini dibersihkan maka data yang berupa kalimat-kalimat ini diklasifikasikan berdasarkan frekuensi kemunculan kata-kata yang sudah dilabeli dari awal. Berikut sampel beberapa kata yang sudah dilabeli. Kata dengan label positif diberikan weight sebesar 1, 0 untuk netral dan -1 untuk kata berlabel negatif

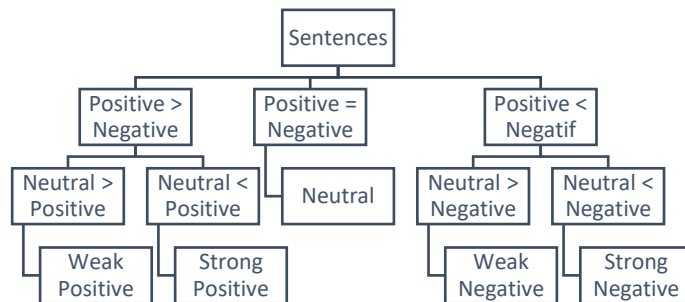
Kata	Label	Weight
Good	Positive	1
Excited	Positive	1
Ordinary	neutral	0
Bad	Negative	-1
Missing	Negative	-1
Boring	Negative	-1
Better	Positive	1
Worse	Negative	-1
Entertaining	Positive	1
Unfortunately	Negative	-1
Fascinating	Positive	1
Not special	Neutral	0
like	Positive	1
Hate	Negative	-1
usual	Neutral	0

C. Decision Tree Classifier

Setelah kita memiliki data training yang sudah dilabeli maka kita dapat menggunakan pohon keputusan untuk mengklasifikasikan kata per kata dalam kalimat ke dalam sebuah kategori yaitu positif, negatif, atau netral dengan pohon keputusan sebagai berikut



Gambar 3.1 Pohon Keputusan Untuk Menentukan Pembobotan sebuah kata



Gambar 3.2 Pohon Keputusan Untuk Menentukan Klasifikasi sebuah kalimat hasil perhitungan

Decision Tree Classifier ini dibentuk dengan menggunakan bantuan algoritma Naïve Bayes untuk menghitung besaran polaritas suatu kalimat berdasarkan frekuensi pembobotan kata-kata yang membangunnya. Naïve Bayes mengelompokkan fitur berdasarkan kecenderungannya untuk muncul dalam sebuah kalimat.

Setiap kata dibobotkan kemudian dicari nilai frekuensi kemunculannya yang kemudian dibandingkan dengan panjang kalimat tersebut

Nilai polaritas suatu kalimat dapat dikalkulasikan dengan rumus sebagai berikut

$$Polarity = \frac{(freq\ pos \times weight) + (freq\ neg \times weight)}{total\ freq}$$

Data Sampel

RT @BORReport: Avengers: Infinity War had the week's most trailer views on YouTube with a massive 77.10M views. Incredibles 2 was second

RT	@BORReport	Avengers	Infinity	War
0	0	0	0	0
had	the	week's	most	trailer
0	0	0	1	0
views	on	YouTube	with	massive
0	0	0	0	1
77.10M	views	Incredibles	2	second
0	0	1	0	0

Maka didapat polaritasnya = ((3 x 1) + (0 x -1)) / 22 = 0.1363

Dari hasil perhitungan polaritas ini dapat disimpulkan bahwa nilai polaritas kalimat ini > 0 sehingga kalimat ini termasuk dalam kategori **Positif**

NME: Did you notice these characters were missing from the new #Avengers trailer?

NME	Did	you	notice	these
0	0	0	0	0
characters	were	missing	from	The
0	0	-1	0	0
New	#Avengers	Trailer	?	
0	0	0	0	

Maka didapat polaritynya = ((0 x 1) + (1 x -1)) / 14 = -0.07

Dari hasil perhitungan polaritas ini dapat disimpulkan bahwa nilai polaritas kalimat ini < 0 sehingga kalimat ini termasuk dalam kategori **Negatif**

D. Hasil Eksperimen

```
C:\Users\Lenovo\LearnPython>py twitterMining.py
What do Twitter say about Avengers Infinity War Trailer ?
Positive tweets percentage: 25 %
Negative tweets percentage: 2 %
Neutral tweets percentage: 72 %
```

Gambar 3.3 Hasil Persentase Analisis sentimen Trailer

Avengers: Infinity War

Setelah program dijalankan, penulis mendapatkan 80 data tweets yang berkaitan atau mengandung query "Trailer Avengers Infinity War" dan diperoleh hasil analisis sentiment seperti gambar di atas. Diperoleh 20 tweet positif, 2 tweet negative dan 58 tweet yang netral. Adapun berikut beberapa hasil dan contoh kalimat yang telah terkategori dari masing-masing kategori

Positive Tweets

Wakanda defends the whole world in the first '#Avengers: #InfinityWar' trailer: <https://t.co/v5J1318W7n>
<https://t.co/gEURxXaITd>

OH MYYY GOOOOODDDDD IM SO EXCITED FOR THIS PLEASE

RT @BORReport: Avengers: Infinity War had the week's most trailer views on YouTube with a massive 77.10M views. Incredibles 2 was second wÔÇª

I liked a @YouTube video <https://t.co/h0qI9TiplK> Marvel Studios' Avengers: Infinity War Official Trailer

Avengers: Infinity War had the week's most trailer views on YouTube with a massive 77.10M views. Incredibles 2 wasÔÇª <https://t.co/tbnyWQd3PM>

Avengers: Infinity War trailer sets record for most views in a single day <https://t.co/BwApGnWilz>

Negative Tweets

NME: Did you notice these characters were missing from the new #Avengers trailer? <https://t.co/ZLTryhW0Qe>
<https://t.co/iKh1D1CqQ4>

Did you notice these characters were missing from the new #Avengers trailer?
<https://t.co/ZPJ5tn03kC> <https://t.co/6XFBhAS8QR>

Neutral Tweets

Avengers: Infinity War' directors hint at trailer release CNET <https://t.co/MSvHxJOj28>

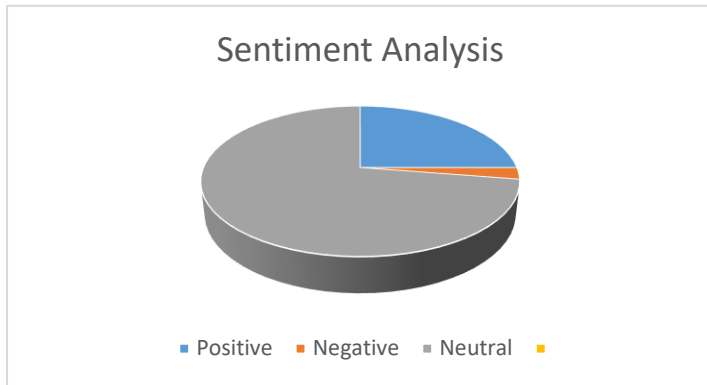
ICYMI: 10 Characters Not In The 'AVENGERS: INFINITY WAR' Trailer <https://t.co/9o8Ihrv4vZ>
<https://t.co/82SvuO9TZL>

RT @FattMernandez: The Avengers Infinity War trailer just beat Justice League at the box office.

RT @capitamrvel: avengers infinity war trailer 90s animated version <https://t.co/KPg7sxbTcR>

RT @TheGeekInsiderM: La Taberna Geek Expr |@s se adentra en el trailer de Avengers: Infinity War. <https://t.co/08duq4dS6K>

E. Data Analysis



Gambar 3.4 Diagram Lingkaran Hasil Klasifikasi Sentimen

Berdasarkan hasil yang diperoleh kita dapat menyimpulkan bahwa sampel respon pengguna sosial media Twitter menunjukkan kecondongan yang masih cenderung netral setelah dirilisnya Trailer Film Avengers ini. Perlu diperhatikan juga dengan jauhnya perbandingan *tweet* positif dan *tweet* negatif maka dapat disimpulkan bahwa trailer ini sudah berhasil menarik perhatian sebagian besar pengguna media sosial Twitter. Hasil analisis ini dapat dijadikan indikator keberhasilan Marvel Studios dalam strategi pemasaran filmnya.

F. Confussion Matrix

Berdasarkan data yang diperoleh kita dapat melakukan validasi model pohon keputusan yang telah kita bangun untuk melakukan klasifikasi dengan menggunakan confusion matrix. Berdasarkan hasil validasi yang dilakukan dari seluruh data diperoleh nilai

Akurasi = 66,25 %
Presisi = 64,8 %
Recall = 66,8 %

		Reality		
		Positive	Neutral	Negative
Test Result	Positive	18	2	0
	Neutral	23	35	0
	Negative	0	2	0

IV. PEMANFAATAN ANALISIS SENTIMEN DALAM INDUSTRI

I. Memetakan Preferensi Publik

Hasil analisis sentiment dapat digunakan untuk memetakan permasalahan-permasalahan yang ada pada konten saat ini. Permasalahan-permasalahan yang muncul ini biasanya terdapat pada sentiment berkategori negative. Pada kategori ini biasanya masyarakat menunjukkan kata kunci yang menjadi permasalahan mengapa sikap mereka terhadap suatu konten menjadi negatif misalnya:

Thanos face looks horrible

Dari hasil evaluasi tersebut dapat dilakukan perbaikan maupun pengembangan konten sehingga permasalahan tersebut dapat diatasi secara tepat sasaran.

Di samping itu kita juga dapat mengetahui secara spesifik hal apa saja yang disukai public sehingga dapat dijadikan sebagai senjata untuk konten yang akan ditampilkan selanjutnya.

II. Komparasi dengan Kompetitor

Analisis Sentimen juga dapat digunakan sebagai komparasi dengan competitor. Contohnya pada kasus ini adalah dengan membandingkan analisis sentimen Trailer Avengers Infinity War dengan Trailer Justice League yang dibuat oleh kompetitor dari Marvel Studios yaitu DC Comics. Dari hasil komparasi dapat diprediksi seberapa besar market share yang bisa diperoleh Marvel Studio dan proyeksi keuntungan dari hasil penayangan film ini kelak.

Berikut contoh data yang bisa dianalisis sebagai komparasi dengan competitor :

RT @FattMernandez: The Avengers Infinity War trailer just beat Justice League at the box office.

V. KESIMPULAN

Konsep Pohon dalam Matematika Diskrit ternyata sangat aplikatif dalam kehidupan sehari-hari. Salah satu pohon yang kegunaannya cukup banyak dan luas pada masa kini adalah pohon keputusan. Pohon keputusan dapat digunakan untuk mengklasifikasikan kecondongan opini masyarakat pada suatu topik spesifik. Kecondongan opini sangatlah penting untuk diperhatikan sebagai indikator keberhasilan suatu produk atau pemasaran suatu produk. Selain itu analisis sentiment juga dapat digunakan sebagai dasar perumusan strategi pengembangan suatu produk.

VI. PENUTUP

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas berkat dan rahmat-Nya penulis dapat menyelesaikan makalah ini dengan baik dan pada waktu yang tepat. Penulis mengucapkan terima kasih kepada kedua orang tua serta teman-teman yang terus memberikan dukungan baik secara moral maupun doa. Ucapan terima kasih turut penulis sampaikan kepada Bapak Dr. Judhi Santoso M.Sc, selaku dosen dari mata kuliah Matematika Diskrit yang telah membimbing penulisan makalah ini. Akhir kata, penulis memohon maaf apabila terdapat kekurangan dan kesalahan kata dalam makalah ini. Penulis berharap makalah ini dapat digunakan sebaik-baiknya dan dikembangkan sehingga lebih menghasilkan manfaat bagi masyarakat luas

REFERENCES

- [1] Jeevanandam J & Dr. S. Koteeswaran, Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis, ARPN Journal of Engineering and Applied Sciences, ISSN 1819-6608, 2015, VOL. 10, NO. 14, 5883 – 5894
- [2] Munir, Rinaldi, Matematika Diskrit, Bandung: Penerbit Informatika Bandung, 2016, ed. 6
- [3] Rosen, Kenneth H., Discrete Mathematics and Its Applications, New York: McGraw-Hill International, 2012, 7th ed.
- [4] <https://celebrity.okezone.com/read/2017/12/03/206/1824246/rekor-demi-rekor-trailer-pertama-avengers-infinity-war-ditonton-230-juta-kali> diakses pada 3 Desember 2017 pada pukul 11.47 WIB
- [5] <http://teknokompas.com/read/2017/07/28/15080057/jumlah-pengguna-mandek-saham-twitter-turun> diakses pada tanggal 3 Desember 2017 pukul 11.28 WIB

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 3 Desember 2017



Christian Wibisono
13516147