

Sentimen Analisis Komentar YouTube Menggunakan

C4.5 Dan Adaptive Boosting

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya*



Oleh :

Redha Bayu Anggara
NIM : 09021381419067

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2019**

LEMBAR PENGESAHAN TUGAS AKHIR

Sentimen Analisis Komentar YouTube Menggunakan C4.5 Dan Adaptive Boosting

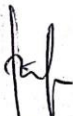
Oleh :

Redha Bayu Anggara
NIM : 09021381419067

Palembang, Juli 2019

Pembimbing I,

Pembimbing II,



Rifkie Primartha, M.T.
NIP. 197706012009121004



If. Muhammad Ihsan Jambak, M. Sc., MM
NIP. 196804052015109101

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004


TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis, 04 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Redha Bayu Anggara
NIM : 09021381419067
Judul : Sentimen Analisis Komentar YouTube Menggunakan C4.5 Dan Adaptive Boosting

1. Pembimbing I

Rifkie Primartha, M.T.
NIP.1971100606740013



.....

2. Pembimbing II

Ir. Muhammad Ihsan Jambak, M. Sc., MM
NIP.198908062015042002



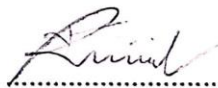
.....

3. Penguji I
Alvi Syahrini Utami, M.Kom
NIP.198603212018032001



.....

4. Penguji II
Mastura Diana Marieska, M.T
NIP. 197812222006042003



.....

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Redha Bayu Anggara
NIM : 09021381419067
Program Studi : Teknik Informatika Bilingual
Judul Skripsi : Sentimen Analisis Komentar YouTube
Menggunakan C4.5 Dan Adaptive
Boosting

Hasil Pengecekan Software *iThenticate/Turnitin* : 14%

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 26 Juli 2019



(Redha Bayu Anggra)

NIM. 09021381419067

KATA PENGANTAR



Alhamdulillah Robbil'Alamin, dengan menyebut nama Allah SWT. Saya berterima kasih atas berkat dan rahmat-Nya berkat petunjuk dan kemudahan sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Dan juga penulis tidak lupa senantiasa mengucapkan syukur bagi Nabi Muhammad SAW berkat tuntunan beliau jua penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun sebagai syarat dalam memenuhi pendidikan program Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini terdapat banyak pihak yang telah memberikan bantuan dan dukungan baik dalam bentuk materil maupun moril. Untuk itu Penulis ingin menyampaikan rasa terima kasih yang tulus kepada :

1. Ayah dan Ibukuku tersayang, H. Yulius Faisol, S.E., M.Si. dan Dra. Lies Narti, adikku Vanisa Larasati dan M.Ikhsan Yudistira seluruh keluarga besarku yang selalu senantiasa mendoakan, menuntun, memberikan motivasi serta dukungan yang luar biasa kepada penulis;
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer;
3. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika dan selaku dosen pembimbing I dan bapak Ir. Muhammad Ihsan Jambak, M. Sc., MM selaku dosen pembimbing II yang telah memberikan arahan, bimbingan, ilmu pengetahuan, nasihat serta memberikan kemudahan bagi penulis dalam proses pengerjaan;

4. Ibu Novi selaku dosen pembimbing saya sebelumnya yang telah memberikan banyak arahan, ilmu pengetahuan serta nasihat sebagai pendorong bagi penulis;
5. Bapak/Ibu ... selaku dosen penguji I dan Bapak/Ibu ... selaku dosen penguji II yang telah memberikan masukan dan ilmu pengetahuan kepada penulis;
6. Bapak Rifkie Primartha, M.T. selaku dosen pembimbing akademik;
7. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika dan staf Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membantu dalam kelancaran penulis selama masa kegiatan perkuliahan;
8. Cynthia Anissa Agatha, partner, sekaligus sahabat yang senantiasa memberikan dukungan serta motivasi yang tak terhingga bagi penulis saat menyelesaikan tugas akhir ini;
9. Keluarga besar Telkomsel Branch Palembang, terkhusus Pak. Dheni Haryanto tempat yang selalu bersedia memberikan masukan dan nasihat untuk penulis dalam semua hal;
10. ICT Semen Baturaja tempat baru untuk belajar yang begitu menyenangkan;
11. Team Marketing Intelligence terimakasih karena setiap hari WhatsApp selalu penuh selalu ada laporan aplikasi gangguan;
12. Sriwijaya Teknologi, Terkhusus kak Ali Bardadi. Yang begitu banyak memberikan *new insight* bagi penulis tentang teknologi informasi dan bisnis;
13. PK-IT start-up asli Palembang, dimana tempat saya belajar tentang memanager waktu;

14. Indofibernet. Farhan, farid dan habib terimakasih untuk segala supportnya kita harus tetap solid Google tidak dibangun dalam semalam;
15. Sahabat-sahabat penulis selama perkuliahan keluarga besar IF Bilingual 2014 yang telah memberikan banyak bantuan, memori dan bekerja sama dengan penulis;
16. Satrio Wijaya, Wahyu Andika, M. Bayu Reksa, Winaldi M Husen, dan Araufu Jevian sebagai tempat berbagi cerita serta pengalaman yang tiada tara bagi penulis;
17. Kakak-kakak tingkat terbaik Fasilkom yang telah banyak memberikan bantuan dan saran selama perkuliahan;
18. BEM KM Fasilkom Unsri Tahun, 2015 (#FasilkomBerdidikasi), 2016 (#KabinetHarmoni), 2017 (#KabinetGaruda) yang telah memberikan kepercayaan bagi penulis untuk berkontribusi selama 3 periode, terima kasih untuk segalanya dan maaf untuk kekeliruan selama menjabat;
19. Rekan-rekan IBGF terbaik yang telah menjadi inspirasi dan memberikan pengalaman baru diluar perkuliahan bagi penulis, terima kasih untuk kebersamaan yang sangat mengesankan bagi penulis;
20. Dulur Real Return: Remin, Kemal, Larry, Doni, Engga, Dicky, Zaky, Nugik, Erlangga dan Tama dulur yang selalu ada dari SMA bagi penulis baik dalam suka, duka dan main gaplek;
21. Teman-teman komunitas dan semua organisasi/komunitas lainnya baik didalam maupun luar kampus yang pernah diikuti penulis, terima kasih;

22. Garbi wadah yang baru-baru ini saya temukan, penuh dengan semangat akan kemajuan dan perubahan;
23. Semua pihak lainnya yang tidak dapat disebutkan satu-persatu yang telah banyak berperan bagi penulis terutama dalam penyelesaian tugas akhir ini, terima kasih banyak atas semuanya.

Dalam proses penyelesaian tugas akhir ini penulis sadar bahwa masih terdapat banyak kekurangan dalam pengetahuan dan ilmu yang terbatas. Penulis akhiri dengan segala hormat, semoga tugas akhir ini dapat memberikan manfaat yang baik bagi kita semua.

Palembang, 26 Juli 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv

MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	xii
DAFTAR GAMBAR	xvi
DAFTAR TABEL	xix

BAB 1 PENDAHULUAN

1.1	Pendahuluan	I-1
1.2	Latar Belakang	I-2
1.3	Rumusan Masalah	I-4
1.4	Tujuan Penelitian	I-5
1.5	Manfaat Penelitian	I-5
1.6	Batasan Masalah	I-5
1.7	Kesimpulan	I-7

BAB II KAJIAN LITERATUR

2.1	Pendahuluan	II-1
2.2	Algoritma Adaptive Boosting	II-1
2.2.1	Pengertian Algoritma Adaptive Boosting	II-1
2.2.2	Struktur Algoritma Adaptive	II-2
2.2.3	Keuntungan Algoritma Adaptive	II-2
2.2.4	Kelemahan Algoritma Adaptive	II-2
2.2.5	Karakteristik Algoritma Adaptive	II-2
2.3	Algoritma C4.5	II-3
2.4	Algoritma C4.5 berbasis Adaptive Boosting	II-4
2.4.1	Langkah Penerapan Adaboost dan C4.5.....	II-5
2.5	<i>Cross Validation</i>	II-5
2.5.1	<i>K-fold cross validation</i>	II-6
2.6	<i>Confusion Matrix</i>	II-6

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan	III-1
-----	-------------------	-------

3.2	Unit Penelitian	III-1
3.3	Data	III-1
3.3.1	Jenis dan Sumber Data	III-1
3.3.2	Sumber Data.....	III-2
3.3.3	Teknik Pengumpulan Data	III-2
3.4	Tahapan Penelitian	III-3
3.5	Diagram Alur Proses Perangkat Lunak	III-7
3.6	Metode Pengembangan Perangkat Lunak	III-15
3.7	Teknik Pengujian.....	III-15
3.8	Penjadwalan Penelitian.....	III-15

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-3
4.2.3	Analisis dan Desain	IV-5
4.2.3.1	Analisis Perangkat Lunak	IV-5
4.2.3.2	Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.3	Analisis Data.....	IV-7
4.2.3.4	Analisis Prapengolahan.....	IV-8
4.2.3.5	Analisis Pembobotan Kata	IV-13
4.2.3.6	Analisis Algoritma C4.5	IV-16
4.2.3.7	Analisis Model Split Validation	IV-24
4.2.3.8	Analisis Metode Adaptive Boosting.....	IV-24
4.2.3.9	Desain Perangkat Lunak.....	IV-26
4.3	Fase Elaborasi	IV-37
4.3.1	Pemodelan Bisnis	IV-37
4.3.1.1	Perancangan Data	IV-38
4.3.1.2	Perancangan Antarmuka	IV-38
4.3.1.3	Kebutuhan Sistem.....	IV-45
4.3.1.4	Sequence Diagram	IV-46
4.4	Fase Konstruksi	IV-52
4.4.1	Kebutuhan Sistem	IV-52
4.4.1.1	Diagram Kelas.....	IV-52
4.4.1.2	Implementasi	IV-53
4.4.1.2.1	Implementasi Kelas.....	IV-54
4.4.1.2.2	Implementasi Antarmuka	IV-55
4.5	Fase Transisi	IV-60
4.5.1	Pemodelan Bisnis	IV-60

4.5.2	Kebutuhan Sistem	IV-60
4.5.3	Rencana Pengujian	IV-61
4.5.3.1	Rencana Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-61
4.5.3.2	Rencana Pengujian <i>Use Case</i> Memuat Data	IV-62
4.5.3.3	Rencana Pengujian <i>Use Case</i> Praproses Data	IV-62
4.5.3.4	Rencana Pengujian <i>Use Case</i> Melihat Data	IV-63
4.5.3.5	Rencana Pengujian Optimasi C4.5 dengan AdaBoost.....	IV-63
4.5.4	Implementasi	IV-64
4.5.4.1	Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-64
4.5.4.2	Pengujian <i>Use Case</i> Memuat Data	IV-66
4.5.4.3	Pengujian <i>Use Case</i> Praproses Data	IV-67
4.5.4.4	Pengujian <i>Use Case</i> Melihat Data	IV-68
4.5.4.5	Pengujian <i>Use Case</i> Optimasi C4.5 dengan AdaBoost	IV-69
4.6	Kesimpulan	IV-70

BAB V HASIL DAN ANALISIS PENELITIAN

5.1	Pendahuluan	V-1
5.2	Data Hasil Uji	V-1
5.2.1	Data <i>Balanced</i>	V-1
5.2.2	Data Tidak <i>Balanced</i>	V-3
5.3	Olah Data	V-3
5.3.1	Uji sama Metode Dengan Data <i>Balanced</i> dan <i>Unbalanced</i>	V-3
5.3.2	Uji Beda Metode dengan Data <i>Balanced</i> dan <i>Unbalanced</i>	V-7
5.4	Analisis Hasil Pengujian.....	V-10
5.4.1	Analisis Peningkatan Akurasi	V-10
5.4.2	Analisis Outliers	V-10
5.4.3	Analisis Performa Klasifikasi	V-12
5.5	Kesimpulan	V-13
5.4.1.	Analisis Peningkatan Akurasi.....	V-10
5.4.2.	Analisis Outliers.....	V-11
5.4.3.	Analisis Performa Klasifikasi.....	V-13
5.6	Kesimpulan	V-13

BAB VI KESIMPULAN DAN SARAN

6.1	Pendahuluan	VI-1
-----	-------------------	------

6.2 Kesimpulan	VI-1
6.3 Saran	VI-2

DAFTAR PUSTAKA	xxi
----------------------	-----

LAMPIRAN	yyy
----------------	-----

DAFTAR GAMBAR

	Halaman
Gambar II-1 Algoritma AdaBoost	II-3
Gambar III-1 Unsupervised Learning	III-5
Gambar III-2 Diagram Alur Proses Umum Perangkat Lunak	III-8
Gambar III-3 Diagram Alur Proses C4.5	III-11
Gambar III-4. Diagram Alur Proses C4.5	III-12
Gambar III-5. Diagram Alur Proses C4.5	III-12
Gambar III-6. Diagram Alur Proses Boosting C4.5	III-14
Gambar IV-1. Algoritma Bubble Sort	IV-17
Gambar IV-2. Pelabelan Kelas	IV-18
Gambar IV-3. Sebelum Dilakukan Unsupervised Missing Value	IV-19
Gambar IV-4. Setelah Dilakukan Unsupervised Missing Value	IV-19
Gambar IV-5. Setelah Dilakukan Unsupervised Discretization	IV-20
Gambar IV-6. Proses Pencarian Root Node	IV-22
Gambar IV-7. Proses Pencarian Leaf Node	IV-23
Gambar IV-8. Penggabungan Kelas Lemah dan Kuat	IV-25
Gambar IV-9. Diagram Use Case Perangkat Lunak	IV-26
Gambar IV-10. Diagram Aktivitas <i>Use Case</i> Klasifikasi dengan C4.5	IV-35
Gambar IV-11. Diagram Aktivitas <i>Use Case</i> Memuat Data	IV-36
Gambar IV-12. Diagram Aktivitas <i>Use Case</i> Praproses Data	IV-36
Gambar IV-13. Diagram Aktivitas <i>Use Case</i> Melihat Data	IV-37
Gambar IV-14. Rancangan Antarmuka Memasukkan Url Video	IV-38
Gambar IV-15. Rancangan Antarmuka Menampilkan Data Komentar	IV-39
Gambar IV-16. Rancangan Antarmuka TFIDF	IV-39
Gambar IV-17. Rancangan Antarmuka Seleksi Atribut	IV-40

Gambar IV-18. Rancangan Antarmuka Pelabelan Kelas	IV-41
Gambar IV-19. Rancangan Antarmuka Hasil Proses Missing Value	IV-41
Gambar IV-20. Rancangan Antarmuka Hasil Proses Diskritisasi	IV-42
Gambar IV-21. Rancangan Antarmuka Pemilihan Klasifikasi	IV-43
Gambar IV-22. Rancangan Antarmuka Hasil Pemilihan Klasifikasi C45	IV-44
Gambar IV-23. Rancangan Antarmuka Hasil Pemilihan Klasifikasi C45 & AdaBoost	IV-44
Gambar IV-24. Diagram Sekuen Hasil Pemilihan	IV-47
Gambar IV-25. Diagram Sekuen Melatih Algoritma C4.5	IV-48
Gambar IV-26. Diagram Sekuen Melihat Data	IV-48
Gambar IV-27. Diagram Sekuen Pra-proses	IV-49
Gambar IV-28. Diagram Sekuen Melihat Data	IV-50
Gambar IV-29. Diagram Sekuen Klasifikasi C45 & AdaBoost	IV-51
Gambar IV-30. Diagram Kelas Perangkat Lunak	IV-53
Gambar IV-31. Memasukan Url Video	IV-55
Gambar IV-32. Tampilan Hasil Spelling Correction	IV-56
Gambar IV-33. Tampilan Hasil TFIDF	IV-56
Gambar IV-34. Tampilan Pemilihan Atribut.....	IV-57
Gambar IV-35 Tampilan Pelabelan Kelas	IV-57
Gambar IV-36 Tampilan Hasil Missing Value	IV-58
Gambar IV-37 Tampilan Hasil Setelah Discretization	IV-58
Gambar IV-38 Tampilan Pemilihan Klasifikasi.....	IV-59
Gambar IV-39 Tampilan Hasil Klasifikasi C4.5.....	IV-59
Gambar IV-40 Tampilan Hasil Klasifikasi C45 & AdaBoost.....	IV-60
Gambar V-1 Visualisasi Sebaran Data Dua Dimensi.....	V-11
Gambar V-2 Visualisasi Sebaran Data Tiga Dimensi	V-11

DAFTAR TABEL

	Halaman
Tabel III-1. Pengembangan Perangkat Lunak dengan Metode RUP	III-16
Tabel III-2 Penjadwalan Penelitian	III-20
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak	IV-4
Tabel IV-2 Kebutuhan Non Fungsional Perangkat Lunak	IV-4
Tabel IV-3. Hasil Tokenizing Contoh Komentar	IV-9
Tabel IV-4 Hasil <i>Spelling Correction</i> Contoh Komentar	IV-10
Tabel IV-5. Hasil <i>Stopword Removal</i> Contoh Komentar	IV-11
Tabel IV-6 Hasil <i>Word Translation</i> Contoh Komentar	IV-11
Tabel IV-7. Hasil <i>Stemming</i> Contoh Komentar	IV-12
Tabel IV-8. Hasil Pembobotan Kata Contoh Komentar	IV-13
Tabel IV-9. Contoh Perangkingan Data	IV-15
Tabel IV-10 Pembobotan Awal	IV-23
Tabel IV-11. Definisi Akor	IV-25
Tabel IV-12. Definisi Use Case	IV-25
Tabel IV-13. Skenario <i>Use Case</i> Klasifikasi C4.5	IV-27
Table IV-14. Skenario <i>Use Case</i> Memuat Data	IV-29
Tabel IV-15. Skenario <i>Use Case</i> Praproses Data	IV-30
Tabel IV-16. Skenario <i>Use Case</i> Melihat Data	IV-31
Tabel IV-17. Skenario <i>Use Case</i> Optimasi C4.5 dengan Adaptive Boosting	IV-32
Tabel IV-18. Implementasi Kelas	IV-52
Tabel IV-19. Rencana Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-60
Tabel IV-20. Rencana Pengujian <i>Use Case</i> Memuat Data	IV-61
Tabel IV-21. Rencana Pengujian <i>Use Case</i> Praproses Data	IV-61

Tabel IV-22. Rencana Pengujian <i>Use Case</i> Melihat Data	IV-62
Tabel IV-23. Rencana Pengujian <i>Use Case</i> Optimasi C4.5 dengan Adaboost	IV-63
Tabel IV-24. Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-64
Tabel IV-25. Pengujian <i>Use Case</i> Memuat Data	IV-65
Tabel IV-26. Pengujian <i>Use Case</i> Praproses Data	IV-67
Tabel IV-27. Pengujian <i>Use Case</i> Melihat Data	IV-68
Tabel IV-28. Pengujian <i>Use Case Weak Learner</i> Menggunakan <i>AdaBoost</i>	IV-69
Tabel V-1. Percobaan dengan <i>data balanced</i>	V-2
Tabel V-2. <i>Precision, recall, dan f-measure data balanced</i>	V-3
Tabel V-3. Percobaan dengan <i>data unbalanced</i>	V-4
Tabel V-4. <i>Precision, recall, dan f-measure data unbalanced</i>	V-5
Tabel V-5 <i>Group Statistics C4.5 balanced dan C4.5 unbalanced</i>	V-6
Tabel V-6 <i>Independent Samples T-test C4.5 balanced dan C4.5 unbalanced</i>	V-7
Tabel V-7. <i>Group Statistics AdaBoost + C4.5 balanced</i> dan <i>AdaBoost + C4.5 Unbalanced</i>	V-8
Tabel V-8 <i>Group Statistics AdaBoost + C4.5 Balanced dan AdaBoost + C4.5 unbalanced</i>	V-9
Tabel V-9 <i>Independent Samples T-test C4.5 balanced dan AdaBoost + C4.5 balanced</i>	V-9
Tabel V-10 <i>Group Statistics C4.5 unbalanced dan AdaBoost + C4.5 unbalanced</i>	V-10
Tabel V-11 <i>Independent Samples T-test C4.5 unbalanced dan AdaBoost + C4.5 unbalanced</i>	

BAB I

PENDAHULUAN

1.1 Pendahuluan

Perkembangan dunia digital telah banyak memberi dampak pada kemudahan mengakses dan berbagi informasi, terutama YouTube semakin hari semakin meningkatnya aktivitas kunjungan masyarakat ke YouTube dengan bermacam cara dalam meresponse suatu video dengan ditonton, like, unlike, berkomentar dan menyebarluaskan. Sehingga video dengan tingkat response dari masyarakat yang tinggi akan menyebabkan *rate* suatu video akan naik kemudian berpeluang menjadi *trending topic*.

Dibalik video *rate* yang naik karena *response* yang tinggi ada satu faktor yang terlupakan yaitu tingkat sentiment komentar masyarakat terhadap video tersebut. Untuk itulah dibutuhkan tambahan informasi lain di dalam menilai tingkat response masyarakat terhadap suatu video, pengenalan jenis sentimen dirasa tepat untuk dijadikan informasi tambahan terhadap suatu video, dimana orientasi

pengenalan jenis sentimen ini nantinya akan berorientasi pada komentar yang ada di YouTube Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017).

1.2 Latar Belakang

Kemampuan untuk menganalisa pengetahuan di dalam teks sehingga teks tersebut menjadi sebuah informasi yang berguna, menjadi tantangan dalam mengetahui jenis sentimen masyarakat didalam sebuah komentar pada video YouTube. Seluruh proses penerapan untuk menemukan pengetahuan pada suatu teks di dalam *computer science* disebut dengan *sentiment analysis* didalam disiplin ilmu *Natural Language Processing*(NLP). Kebutuhan untuk memahami informasi didalam kumpulan teks yang ada, saat ini terjadi diseluruh bidang baik bisnis, sains dan teknologi. Sehingga ketersediaan tools untuk membantu organisasi dalam menganalisa tingkat sentimen sangatlah dibutuhkan. Menganalisa suatu sentimen komentar YouTube sangatlah mungkin dilakukan, tetapi terkadang data teks yang ada banyak dan bermacam bentuk dan dapat menimbulkan ambiguitas sehingga sulit untuk dikenali jenis sentimennya.

Analisis sentimen merupakan bidang komputasi yang mempelajari suatu opini, sikap, atau emosi dari suatu entitas. Suatu entitas bisa saja berupa individu, topik, maupun kejadian. Dalam analisis sentimen kumpulan entitas dibagi menjadi beberapa kelas-kelas sentimen, misalkan positif, negatif, atau netral. Salah satu metode pohon keputusan yang dapat digunakan pada analisis sentimen sebagai acuan untuk mengklasifikasi tingkat sentimen adalah Algoritma C4.5 merupakan

algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang efisien dalam menangani atribut bertipe diskret atau numerik Hermanto, B., Sn, A., & Putra, F. P. (2017). Dalam mengkonstruksi pohon keputusan, Algoritma C4.5 langsung mengekstraksi seluruh sample data latih dan mengkonstruksikannya menjadi model *decision tree* berdasarkan perhitungan Gain Ratio. Hal inilah yang menjadi salah satu kelemahan dari algoritma C4.5 dalam mengkonstruksi sebuah *decision* tanpa mengidentifikasi dan melihat hubungan terhadap faktor-faktor yang mempengaruhi data latih Bisri & Satria Wahono (2015)., sehingga akan berakibat pada desain *decision tree* yang kurang optimal.

Untuk itulah dibutuhkan sebuah metode tambahan untuk mempelajari keterkaitan apa saja yang mempengaruhi data latih, salah satunya algoritma Adaptive Boosting adalah metode *ensemble learning* dengan menggunakan *loss function* fungsi *exponential* untuk mempelajari setiap kontradiksi pada data training, yang berguna untuk *feature selection* dengan cara kerja membandingkan dan menggabungkan populasi kelas lemah dan kuat dalam sub data training sebelum akhirnya dilakukan pembentukan pohon keputusan.

Adaptive Boosting merupakan suatu prosedur yang berulang untuk mengubah distribusi data latih agar model yang terbentuk dapat berkonsentrasi pada data yang sulit untuk diklasifikasi Parag, Porikli, & Elgammal (2008). C4.5 dengan *boosting* akan membentuk prosedur berulang dengan cara melakukan pemilihan *sub data training* pada setiap iterasi dimana data yang sulit diprediksi pada iterasi

sebelumnya akan dimasukkan ke dalam distribusi data iterasi berikutnya. Proses konstruksi pohon keputusan akan selesai apabila pada iterasi *ke-i* diketahui atribut telah habis atau *error* bernilai lebih besar dari 0.5 .

Penelitian oleh Lila Dini Utami. (2015) menerapkan metode *AdaBoost* dan *Naive Bayes* untuk melakukan proses analisis sentimen *review* Restoran mendapatkan akurasi model sebesar 99.50% dibandingkan 70.00% dengan hanya menerapkan *Naive Bayes*. Hasil penelitian ini menunjukkan bahwa dengan adanya kombinasi dari *AdaBoost* mampu meningkatkan akurasi dari *Naive Bayes*.

Berdasarkan hal tersebut, maka penelitian ini akan menguji perbandingan hasil akurasi yang dihasilkan oleh C4.5 dan *Adaboost* dan C4.5 tanpa *Adaboost* dalam mengklasifikasi sentimen komentar video YouTube, *Adaboost* merupakan *ensemble learning* yang sering digunakan pada algoritma *boosting*. *Boosting* bisa dikombinasikan dengan *classifier* algoritma yang lain untuk meningkatkan performa klasifikasi. Tentunya secara intuitif, penggabungan beberapa model akan membantu jika model tersebut berbeda satu sama lain. *Adaboost* dan variannya telah sukses diterapkan pada beberapa bidang (*domain*) karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar. (Eka Listiana, 2017).

1.3 Rumusan Masalah

Masalah yang dirumuskan pada penelitian ini adalah apakah Adaptive Boosting (*AdaBoost*) dapat meningkatkan akurasi algoritma C4.5 pada analisis

sentimen. Untuk menyelesaikan masalah di atas, pada penelitian ini dibagi menjadi beberapa *research question*, sebagai berikut:

1. Bagaimana mekanisme penerapan algoritma Adaptive Boosting dalam mengubah distribusi data latih agar model yang terbentuk dapat berkonsentrasi pada data-data yang sulit untuk diklasifikasi.
2. Apakah algoritma Adaptive Boosting dapat meningkatkan akurasi C4.5 dalam klasifikasi jenis sentiment.

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut:

1. Menerapkan dan Menganalisis kombinasi algoritma C4.5 dan Adaptive Boosting pada analisis sentimen komentar YouTube.
2. Mengembangkan perangkat lunak yang dapat mengklasifikasikan jenis sentimen komentar YouTube menggunakan metode C4.5 dan AdaBoost.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Hasil penelitian dapat digunakan untuk mengetahui akurasi dari algoritma C4.5 dengan AdaBoost dan tanpa AdaBoost.
2. Hasil penelitian dapat digunakan untuk seseorang mengetahui jenis-jenis sentimen di video youtube.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini mengklasifikasi sentimen komentar YouTube menjadi 3 kategori netral, positif dan negatif.
2. Komentar yang dianalisis dan klasifikasi adalah komentar yang menggunakan Bahasa Indonesia.
3. Dataset yang digunakan adalah data komentar youtube berbahasa Indonesia yang diperoleh dari Application Programming Interface (API) yang tersedia di situs Google Developer.
4. Format dataset yang digunakan berbentuk Javascript Orientation Nation (JSON).

- Methods, D. P. (2011). Data Preprocessing Techniques for Data Mining. *Science*.
<https://doi.org/10.1002/pmj.20069>
- Gashler, M. S., Smith, M. R., Morris, R., & Martinez, T. (2016). Missing value imputation with unsupervised backpropagation. *Computational Intelligence*.
<https://doi.org/10.1111/coin.12048>
- Asian, J., Williams, H. E., & Tahaghoghi, S. M. M. (2005). Stemming Indonesian. In *Conferences in Research and Practice in Information Technology Series*.
<https://doi.org/10.1145/1316457.1316459>
- Lakshminarayan, K., Harp, S. A., Goldman, R., & Samad, T. (1996). Imputation of Missing Data Using Machine Learning Techniques. *KDD-96 Proceedings*.
<https://doi.org/10.1016/j.inffus.2012.08.007>
- Parag, T., Porikli, F., & Elgammal, A. (2008). Boosting adaptive linear weak classifiers for online learning and tracking. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*.
<https://doi.org/10.1109/CVPR.2008.4587556>
- Das, A. (2017). Sentiment analysis. In *8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017*.
<https://doi.org/10.1109/ICCCNT.2017.8203921>
- Bisri, A., & Satria Wahono, R. (2015). Penerapan Adaboost untuk Penyelesaian Ketidakseimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode Decision Tree. *Journal of Intelligent Systems*.
- Hermanto, B., Sn, A., & Putra, F. P. (2017). Analisis Kinerja Decision Tree C4.5 Dalam Prediksi Potensi Pelunasan Kredit Calon Debitur. *Inovtek Polbeng Seri Informatika*.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*.
<https://doi.org/10.15575/join.v2i1.71>
- Ramos, J. (2003). *Using TF-IDF to Determine Word Relevance in Document Queries. Processing*. <https://doi.org/10.1.1.121.1424>
- Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017). Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine. *International Journal of Computer Science and Software Engineering ISSN*.

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and Unsupervised Discretization of Continuous Features. In *Machine Learning Proceedings 1995*. <https://doi.org/10.1016/B978-1-55860-377-6.50032-3>

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.279>