

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
TERHADAP DEBAT PEMILIHAN PRESIDEN 2019
MENGUNAKAN *METODE NAÏVE BAYES CLASSIFIER***

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat

Sarjana Teknik Informatika



Dibuat Oleh:

ROBERT ADVENTO BEDING

14 07 07880

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
YOGYAKARTA
2019**

LEMBAR PENGESAHAN

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP DEBAT PEMILIHAN PRESIDEN 2019 MENGGUNAKAN *METODE NAÏVE BAYES CLASSIFIER*

Yogyakarta, November 2019

Robert Advento Beding
140707880

Menyetujui,

Pembimbing I



Martinus Maslim, S.T., M.T.

Pembimbing II



Paulus Mudjihartono, S.T., M.T.

Penguji I



Martinus Maslim S.T., M.T.

Penguji II



Ir.A. Djoko Budiyanto, M.Eng., Ph.D.

Penguji III



Th. Adi Purnomo Sidhi, S.T., M.T.

Mengetahui,

Dekan Fakultas Teknologi Industri




Dr. Drs. A. Teguh Siswanto, M.Sc.

KATA PENGANTAR

Segala puji dan syukur penulis haturkan kepada Petara Yesus Kristus, berkat anugerah-Nya yang melimpah, serta kemurahan dan kasih setia yang selalu menyertai sehingga penulis dapat menyelesaikan laporan skripsi dengan baik dan lancar. Tugas akhir merupakan salah satu syarat untuk mencapai derajat sarjana dari Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta. Penulis menyadari bahwa dalam pembuatan tugas akhir, penulis telah mendapatkan bantuan, bimbingan, dan dukungan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Petara (Tuhan Yang Maha Esa) Yesus Kristus, yang selalu melimpahkan anugerah, kemurahan dan kasih setia-Nya.
2. Bapak Martinus Maslim, S.T., M.T., selaku Ketua Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
3. Bapak Martinus Maslim, S.T., M.T., selaku Dosen Pembimbing I yang telah bersedia membimbing, selalu memberikan solusi dan gagasan ide kepada penulis sehingga skripsi ini dapat diselesaikan dengan baik.
4. Bapak Paulus Mudjihartono, Ph.D selaku Dosen Pembimbing II yang telah bersedia membimbing, selalu memberikan solusi dan gagasan ide kepada penulis sehingga skripsi ini dapat diselesaikan dengan baik.
5. Bapak Prof. Suyoto, Ph.D selaku dosen pembimbing akademik dari semester awal sampai akhir yang selalu memberikan solusi dan motivasi sehingga selama perkuliahan dapat berjalan dengan baik.
6. Seluruh Dosen, Staf Pengajar, dan Karyawan Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama masa kuliah di Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
7. Keluarga tercinta, ibu, Ayah, Abang, Adik serta saudara yang lain, yang tidak pernah lupa untuk selalu memberi dukungan doa, semangat, dan

motivasi kepada penulis agar dapat menyelesaikan skripsi dengan tepat waktu.

8. Teman seperjuangan skripsi, yang tidak mungkin disebutkan satu persatu selama pembuatan skripsi yang selalu membantu dan mendukung penulis dalam menyelesaikan kripsi.
9. Teman-teman dari sukses, gaje group dan kantuy yang tidak mungkin disebutkan satu persatu yang selalu memberi dukungan dan semangat kepada penulis agar dapat menyelesaikan skripsi dengan tepat waktu.
10. Teman-teman seperantauan KBMDA Yogyakarta yang secara tidak langsung memberikan semangat dan dukungan sehingga penulis dapat menyelesaikan Tugas Akhir ini.
11. Teman-teman, yang selalu mendukung dan memberi semangat serta menjadi motivasi penulis dalam menyelesaikan skripsi ini.

Yogyakarta, November 2019

Robert Advento Beding
140707880

DAFTAR ISI

JUDUL.....	i
LEMBAR PENGESAHAN	ii
KATA PENGANTAR	iii
Daftar Isi	V
Daftar Gambar.....	Vii
Daftar Tabel	viii
Intisari	ix
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Metodologi Penelitian	3
1.6. Sistem Penulisan Tugas Akhir	Error! Bookmark not defined.
BAB II.....	7
TINJAUAN PUSTAKA	7
BAB III	12
LANDASAN TEORI.....	12
3.1. Twiter	12
3.2. Data Mining	13
3.3. Analisis Sentiment	13
3.4. Metode Naïve Bayes Classifier.....	14
3.5. Pengumpulan Data	15
3.6. <i>Preprocessing atau Cleaning</i>	16
3.7. WEKA	16
3.8. <i>K-fold Cross Validation</i>	17
3.9. <i>Confusion Matrix</i>	17
3.10. <i>Zero-R</i>	18

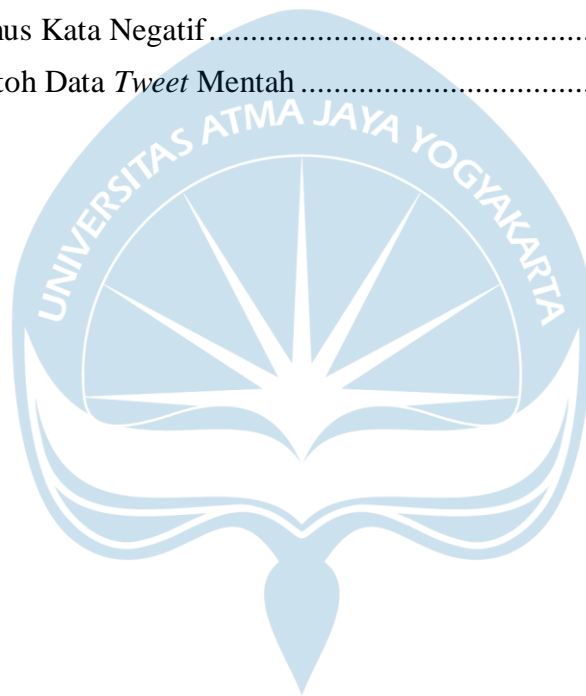
3.11. <i>One-R</i>	18
3.12. <i>N-Gram</i>	19
BAB IV	20
METODOLOGI PENELITIAN	20
4.1. Proses Pengumpulan Data	20
4.2. Pre-Processing Data	21
4.3. Labeling Data.....	23
4.4. Evaluasi Model	32
4.5. Perbandingan Naïve Bayes Classifier dengan metode klasifikasi lain ..	33
4.6. Interpretasi Data.....	33
BAB V	35
HASIL DAN PEMBAHASAN	35
5.1. Data	35
5.2. <i>Preprocessing Data</i>	37
5.3. Penetapan Kelas Sentimen	39
5.4. Validasi Metode <i>Naïve Bayes Classifier</i>	40
5.5. Validasi Metode <i>Zero-R</i>	44
5.6. Validasi Metode <i>One-R</i>	46
5.7. Perbandingan antara Metode <i>Naïve Bayes Classifier</i> dengan <i>Zero-R</i> dan <i>One-R</i>	48
5.6. Interpretasi Data <i>Tweet</i> Debat Pilpres 2019	49
BAB VI	52
KESIMPULAN DAN SARAN	52
6.1. Kesimpulan.....	52
6.2. Saran.....	52
Daftar Pustaka	54
LAMPIRAN	5Error! Bookmark not defined.

Daftar Gambar

Gambar 3.1 Rumus Dasar Dalam Metode Naïve Bayes Classifier	15
Gambar 3.2 Tabel <i>Confusion Matrix</i>	17
Gambar 4.1 Alur Pengerjaan Analisis Sentimen.....	20
Gambar 4.2 Alur Pengambilan Data Dari Twitter.....	21
Gambar 4.3 Ilustrasi Proses Stemming.....	23
Gambar 4.4 Contoh Hasil Dari Kolom Positif	27
Gambar 4.5 Contoh Hasil Dari Kolom Negatif.....	28
Gambar 4.6 Contoh Perhitungan Skor Positif dan Negatif.....	29
Gambar 4.7 Contoh Perhitungan Kelas Positif, Negatif dan Netral.....	29
Gambar 4.8 Contoh Jumlah Data Positif	30
Gambar 4.9 Contoh Jumlah Data Negatif.....	30
Gambar 4.10 Contoh Jumlah Data Netral.....	31
Gambar 4.11 Contoh Jumlah Total Data	31
Gambar 4.12 Contoh Diagram Sentimen.....	31
Gambar 4.13 Ilustrasi <i>K-fold Cross Validation</i>	32
Gambar 4.14 Visualisasi <i>Confusion Matrix</i>	33
Gambar 4.15 Visualisasi N-gram	34
Gambar 5.1 Grafik <i>Tweet</i> Berdasarkan Kombinasi <i>Keyword</i>	35
Gambar 5.2 Grafik Kelas Sentimen.....	40
Gambar 5.3 Diagram Garis Dari Metode <i>Naïve Bayes Classifier</i>	41
Gambar 5.4 Hasil Curve ROC (Receiver Operating Characteristic) Positif Menggunakan 20-fold Cross Validation.....	42
Gambar 5.5 Hasil Curve ROC (Receiver Operating Characteristic) Negatif Menggunakan 20-fold Cross Validation.....	43
Gambar 5.6 Hasil Curve ROC (Receiver Operating Characteristic) Netral Menggunakan 20-fold Cross Validation.....	43
Gambar 5.7 Diagram Garis Dari Metode Zero-R.....	45
Gambar 5.8 Persentase Dari Data <i>Tweet</i> Debat Pilpres 2019	49
Gambar 5.9 Diagram Frekuensi Kata Pada Debat Pilpres 2019	51

Daftar Tabel

Tabel 2.1 Perbandingan Antar Peneliti Terdahulu	10
Tabel 4.1 Ilustrasi Proses Celansing	22
Tabel 4.2 Ilustrasi Proses Tokenisasi.....	22
Tabel 4.3 Ilustrasi Proses <i>Stopwords Removal</i>	23
Tabel 4.4 <i>Scoring Sentiment</i>	24
Tabel 4.5 Kamus Kata Positif	25
Tabel 4.6 Kamus Kata Negatif.....	25
Tabel 5.1 Contoh Data <i>Tweet</i> Mentah	36



Intisari

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP DEBAT PEMILIHAN PRESIDEN 2019 MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Intisari

Robert Advento Beding

14 07 07880

Analisis sentimen merupakan salah satu cara untuk memprediksi sebuah sasaran, dengan melakukan penambangan data yang merupakan kumpulan data mentah dan diolah menjadi data yang siap dianalisis. Sosial media yang digunakan dalam melakukan analisis adalah *twitter*, dengan kemudahan dalam penggunaannya dan dengan pertimbangan data yang beragam. Opini masyarakat tentang terselenggaranya debat pilpres 2019 di Indonesia merupakan salah satu cara agar dapat mengetahui bagaimana pendapat dan respon terhadap acara tersebut. Dengan melakukan analisis sentimen dengan melihat pandangan masyarakat apakah menarah pada sifat yang positif, netral atau negatif.

Dari data yang berisi tentang opini masyarakat pada debat pilpres 2019, metode yang digunakan dalam melakukan analisis adalah metode *Naïve Bayes Classifier* dapat membantu dalam menganalisis, dengan pertimbangan tingkat akurasi yang tinggi dan lebih sederhana. Dengan melakukan perbandingan antara metode *Naïve Bayes Classifier*, metode *One-R* dan *Zero-R*.

Hasil pengujian terhadap data *tweet* Debat Pilpres 2019 menunjukkan bahwa metode *Naïve Bayes Classifier* dapat menghasilkan akurasi yang lebih baik dari metode *One-R* dan *Zero-R*. Dengan nilai pada proses *20-k fold cross validation* dengan rata-rata akurasi berjumlah 98.633%, *precision* 0.986%, *recall* 0.987%, *F-Measure* 0.987% dari 45.938 data *tweet* yang dimiliki. Dan kata-kata yang paling sering digunakan pada *tweet* debat Pilpres 2019 dengan

menggunakan metode *N-gram*, adalah gagpunicorn, debatcawapres, debatpilpres, jokowi dan debat.

Kata kunci : Analisis sentimen, Data mining, Twitter, Debat Pilpres 2019, Naïve bayes classifier, N-gram.

