

Analisis Sentimen Terhadap Universitas Gunadarma Berdasarkan Opini Pengguna Twitter Menggunakan Metode *Naive Bayes Classifier*

Imam Mulya dan Cut Maisyarah Karyati

Universitas Gunadarma, Indonesia
E-mail : imammulya21@gmail.com, csyarah@gmail.com

Abstrak

Dalam upaya meningkatkan peminat mahasiswa baru, Universitas Gunadarma perlu mengetahui dan memperoleh opini masyarakat mengenai reputasi dan layanannya sendiri. Hal ini dapat dianalisa melalui kalimat pada media sosial diantaranya melalui tweet pada twitter. Data yang digunakan pada penelitian ini berasal dari twitter yang berupa data opini pengguna komentar dari situs: <https://www.twitter.com> dengan kata kunci Gunadarma pada kolom pencarian, data opini yang diambil 365 data pada tanggal 26 April 2020 yang didapatkan dengan menggunakan bantuan aplikasi Crawling tweet Twitter pada sistus: <https://netlytic.org/> yang di generated otomatis ke dalam format .csv. Data latih dianalisis berdasarkan nilai atributnya dengan menggunakan suatu metode klasifikasi. Proporsi data latih berjumlah 187 data dari total 267 data. Sementara jumlah data uji adalah sebanyak 80 data dari total data. Penelitian ini menggunakan bahasa R dengan IDE RStudio. Kemudian proses klasifikasi dilakukan menggunakan 2 metode yakni: *lexicon based* untuk menentukan kalimat supaya memiliki makna sentimen positif atau negatif dan *naive bayes classifier* untuk menghitung akurasi antara prediksi dan hasil actual dari sistem. Analisis sentimen memperoleh hasil 211 sentimen positif dan 56 sentimen negatif. Hasil analisis ini divisualisasikan dalam bentuk histogram, diagram pie dan *wordcloud*. Pada penelitian analisis sentimen ini didapatkan hasil akurasi klasifikasi data tweet dari sistem analisis sentimen dengan menggunakan naive bayes classifier sebesar 86.42% dengan kesalahan system sebesar 13.58%. Berdasarkan hasil analisis sentimen tersebut dapat disimpulkan bahwa sentimen pengguna Twitter terhadap Universitas Gunadarma memiliki kecenderungan sentimen positif. Analisis sentimen terhadap Universitas Gunadarma dapat dimanfaatkan pihak Universitas Gunadarma dan dijadikan acuan untuk kemajuan dan meningkatkan kinerja dari berbagai pihak terkait.

Kata kunci : Analisis Sentimen, Kalimat Positif, Kalimat Negatif, Naive Bayes, Universitas Gunadarma

Pendahuluan

Teknologi informasi dapat meningkatkan kinerja dan memungkinkan berbagai kegiatan dapat dilaksanakan dengan cepat, tepat dan akurat, sehingga dapat memberikan kemajuan yang signifikan dalam berbagai bidang

seperti bermunculannya berbagai jenis kegiatan yang berbasis teknologi informasi, aplikasi atau perangkat lunak pada beberapa jenis kebutuhan seperti media sosial. Media sosial adalah sebuah media online, dimana para penggunaanya bisa dengan mudah berpartisipasi, berbagi dan menciptakan isi meliputi blog, je-

jaring sosial, forum dan dunia virtual Blog jejaring social serta wiki yang merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia [1].

Pemanfaatan media sosial tidak hanya untuk berkomunikasi antara teman saja, tetapi sering juga dijadikan sebuah sarana untuk menyampaikan suatu aspirasi bagi masyarakat khususnya masyarakat Indonesia mengenai masalah pemerintah, ataupun masalah yang berhubungan dengan dunia pendidikan. Salah satu aspirasi yang disampaikan ini seperti beberapa hastag mengenai Universitas Gunadarma yang berhubungan dengan sentimen negatif dan positif. Dalam upaya meningkatkan peminat mahasiswa baru, Universitas Gunadarma perlu mengetahui dan memperoleh opini masyarakat mengenai reputasi dan layanannya sendiri, karena apabila terdapat lebih banyak sentimen negatif daripada sentimen positif maka dapat membuat reputasi Universitas Gunadarma menjadi buruk di mata masyarakat sehingga mengurangi minat calon mahasiswa yang ingin bergabung dengan Universitas Gunadarma. Permasalahan ini diangkat peneliti dengan membuat sebuah analisis hastag terkait sentimen apa yang terjadi pada masyarakat baik sentimen positif maupun sentimen negatif. Analisis sentimen ini menggunakan metode *Naive Bayes Classifier* yang merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas dasar [2]. Adapun alasan peneliti melakukan analisis sentimen terhadap Universitas Gunadarma adalah agar dapat dimanfaatkan pihak Universitas Gunadarma dan dijadikan acuan untuk kemajuan dan meningkatkan kinerja dari berbagai pihak terkait.

Analisis sentimen pada penelitian ini menggunakan data dari twitter yang merupakan layanan jejaring sosial dan microblog daring yang memungkinkan penggunaannya untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan kicauan (tweet) [3].

Penelitian tentang analisis sentimen telah banyak dilakukan dengan menggunakan metode Naive Bayes dari Tweet Twitter yang menghasilkan nilai akurasi lebih besar dari 90% [4, 5]

Berdasarkan uraian diatas peneliti tertarik melakukan analisis sentimen pengguna Twitter

terhadap Universitas Gunadarma berdasarkan opini dari tweet Twitter menggunakan metode *Naive Bayes Classifier*. Alasan pemilihan objek penelitian ini karena pada saat tahun ajaran baru ini, universitas sedang menjaring calon mahasiswa dan salah satunya melalui media sosial Twitter. Data sentimen digunakan untuk melihat tweet pengguna Twitter terhadap Universitas Gunadarma tentang apa yang mereka bicarakan, positif atau negatif. Universitas Gunadarma merupakan salah satu Universitas terbaik di Indonesia dan memiliki banyak ulasan atau opini di media sosial salah satunya di Twitter.

Text Mining

Teknik yang dapat digunakan untuk melakukan klasifikasi text adalah Text Mining yang merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [6]. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada text mining adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Tugas khususnya adalah pengkategorisasian teks dan pengelompokan teks. Text mining lebih luas dapat diartikan sebagai proses mencari tahu secara intensif dimana pengguna berinteraksi dengan kumpulan dokumen yang merupakan sumber data pada text mining dan pola yang menarik tidak ditemukan pada record database yang terbentuk melainkan pada data kata per kata yang tidak terstruktur pada kumpulan dokumen [7]. Pada dasarnya proses kerja dari text mining banyak mengadopsi dari penelitian data mining namun yang menjadi perbedaan adalah pola yang digunakan oleh text mining diambil dari sekumpulan bahasa alami yang tidak terstruktur sedangkan dalam data mining pola yang diambil dari database yang terstruktur [8].

Analisis Sentimen

Opini mining (analisis sentimen) merupakan salah satu cabang penelitian Teks Mining. Opini mining adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Jika diberikan suatu set dokumen teks yang berisi opini mengenai suatu objek, maka opinion mining bertujuan untuk

mengekstrak atribut dan komponen dari objek yang telah didokumentasi pada setiap dokumen dan untuk menentukan apakah komentar tersebut bermakna positif atau negatif [9]. Analisis Sentimen digunakan untuk mengotomatiskan proses identifikasi pendapat apakah itu adalah pandangan positif atau negatif. Sistem ini dapat mengekstrak opini publik tentang topik tertentu, produk atau jasa yang tertanam dalam teks yang tidak terstruktur [10]. Analisis sentimen dapat dibedakan berdasarkan sumber datanya, beberapa level yang sering digunakan dalam penelitiannya adalah pada level dokumen dan pada level kalimat. Berdasarkan level sumber datanya analisis sentimen terbagi menjadi 2 kelompok besar yaitu: coarse-grained dan ned-grained. Pada analisis sentimen coarse-grained dilakukan pada level dokumen. Secara garis besar fokus utama dari jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah sentimen positif atau sentimen negatif. Ned-grained dilakukan pada level kalimat. Fokus utamanya adalah menentukan sentimen pada setiap kalimat [11].

Naive Bayes

Naive bayes yaitu teorema dalam statistika untuk menghitung peluang. *Bayes Optimal Classifier* menghitung peluang dari satu kelas dari setiap kelompok atribut yang ada dan menentukan kelas mana yang paling optimal. Adapun kelebihan dari penggunaan *Naive Bayes Classifier* dalam klasifikasi dokumen dapat ditinjau dari prosesnya yang mengambil aksi berdasarkan data yang telah ada sebelumnya [12]. Oleh karena itu, klasifikasi dokumen dengan metode ini dapat dipersonalisasi, maksudnya adalah proses klasifikasi dokumen dapat disesuaikan dengan sifat dan kebutuhan, informasi ini dapat dilihat secara nyata melalui hashtag filtering, dengan mengetikkan nama filter hashtagnya. Proses klasifikasi ini menggunakan *Naive Bayes classifier* pengklasifikasi hashtag otomatis ini dapat disesuaikan sehingga meminimalisasi aksi salah pengklasifikasi secara personal dengan memperoleh informasi sentimen positif atau negatif.

Pengumpulan Data

Data opini pengguna komentar diambil dari situs: <https://www.twitter.com> dengan kata kunci Gunadarma pada kolom pencarian, data

opini yang diambil 365 data pada tanggal 26 April 2020 yang didapatkan dengan menggunakan bantuan aplikasi Crawling tweet Twitter pada situs: <https://netlytic.org/> yang di generated otomatis ke dalam format .csv. Data yang digunakan adalah data tweet pengguna Twitter. Berikut ini salah satu contoh tweet pengguna Twitter yang memiliki makna sentimen positif: “Universitas Gunadarma terbuka luas”. Sedangkan ini contoh tweet pengguna Twitter yang memiliki makna sentimen negatif: “Universitas Gunadarma masalah yang di demo kemarin sudah beres kah atau belum mau masuk disitu labil banget”. Data masukan untuk analisis sentimen pada penelitian ini merupakan kumpulan opini yang sudah dikategorikan kedalam data uji ataupun data latih. Data opini terdiri dari 187 data latih dan 80 data uji.

Data Latih

Dalam analisis sentimen, data latih digunakan pada tahap pelatihan. Data latih di analisis berdasarkan nilai nilai atributnya dengan menggunakan suatu metode klasifikasi. Proses pembelajaran untuk melatih sistem agar dapat mengklasifikasikan suatu kata apakah termasuk dalam sentimen positif atau sentimen negatif pada saat proses atau tahap pengujian. Proporsi data latih jumlahnya lebih banyak dibandingkan dengan data uji. Dari total data 267 data, jumlah data latih dalam penelitian ini sebanyak 187 data.

Data Uji

Data uji adalah sekumpulan data yang merupakan proses ini analisis sentimen yakni proses pengujian menggunakan metode *Naive Bayes Classifier*. Dari hasil pengujian dengan data uji didapatkan akurasi sistem yang dibandingkan dengan pengklasifikasian dengan metode lexicon based. Proporsi data uji jumlahnya lebih sedikit dibandingkan dengan data latih. Dari total 267 data, jumlah data uji pada penelitian ini sebanyak 80 data.

Analisis Sentimen terhadap Universitas Gunadarma

Dalam tahap analisis sentimen, struktur data yang baik dapat memudahkan proses analisis sentimen secara otomatis. Oleh karena

itu, data opini yang strukturnya masih kurang terstruktur diubah menjadi lebih terstruktur sesuai dengan kebutuhan. Tahap untuk melakukan analisis sentimen dapat dilihat pada Gambar 1.



Gambar 1: Alur Analisis Sentimen

Tahap awal analisis sentimen adalah membaca data opini (dataset) yang merupakan data opini pengguna twitter terhadap Universitas Gunadarma. Selanjutnya dilakukan tahap pembersihan (pre-processing) yaitu mengubah kapitalisasi, menghilangkan tanda baca / pembatas (delimiter), menghilangkan nama username pengguna Twitter seperti (@), dan menghilangkan Url yang terdapat pada tweet seperti (http//). Dataset yang sudah melewati tahap pembersihan (pre-processing) dilakukan klasifikasi menggunakan metode lexicon based untuk menentukan sentimen data positif atau negatif. Setelah itu, proses pelabelan data latih dan data uji dengan proporsi 70:30 yang akan digunakan pada tahap klasifikasi menggunakan metode Naïve Bayes Classifier, terlebih dulu dilakukan proses pembelajaran menggunakan data latih, kemudian dilakukan proses pengujian menggunakan data uji untuk mendapatkan hasil klasifikasi dan melihat akurasi mengu-

nakan perhitungan confusion matrix dari klasifikasi yang dilakukan oleh sistem.

Tahap Membaca Data Opini

Tahap membaca data komentar merupakan tahap utama yang dilakukan sebelum ke tahap selanjutnya. File data opini (dataset) harus didefinisikan terlebih dahulu agar dapat di proses. Berikut penggalan sintaks untuk mendefinisikan file data opini (dataset) untuk tahap pembersihan (pre-processing).

```

DatasetGunadarma
<read.csv("D:\tesis\Gunadarma_Dataset_clean.csv")
DatasetGunadarma$text <- as.factor(DatasetGunadarma$text)
    
```

Tahap Pembersihan (Pre-processing)

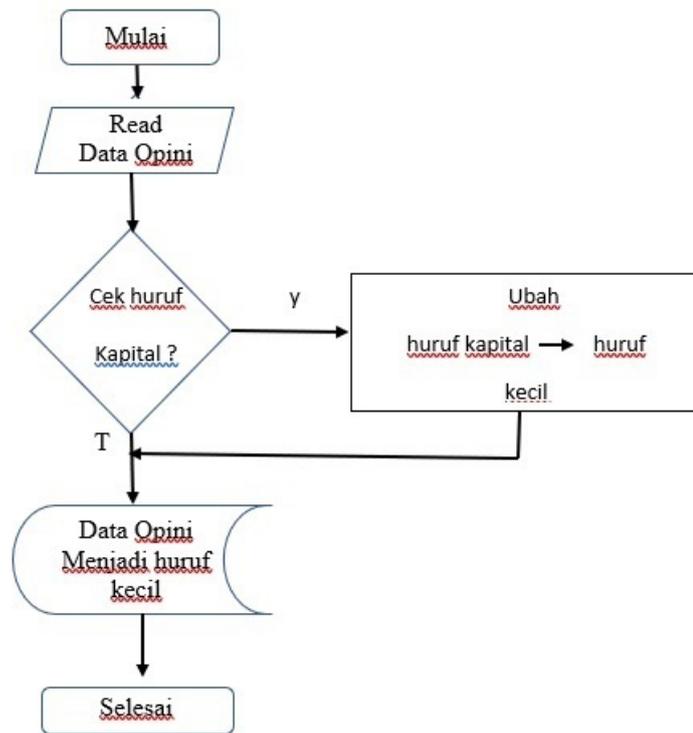
Tahap pembersihan atau pre-processing merupakan tahap yang dilakukan sebelum ke tahap metode klasifikasi. Tujuan dari tahap ini adalah menyeragamkan ukuran kata, menghilangkan username pengguna Twitter. Tahap pembersihan (pre-processing) meliputi mengubah kapitalisasi, menghapus tanda baca, menghapus Url, menghapus username pengguna Twitter. Dalam tahap pre-processing dibutuhkan package R yang harus diinstall dan dipanggil pada saat program dijalankan di RStudio yaitu:

```

install.packages("tm")
install.packages("corpus")
install.packages("plyr")
install.packages("stringr")
install.packages("wordcloud")
library("tm") #untuk proses text mining
library("corpus") #untuk membaca data dari
suatu file
library("plyr") #untuk split data
library("stringr") #untuk proses data string
library("wordcloud") #untuk pembuatan word
cloud
    
```

Mengubah Kapitalisasi

Dalam tahap ini dilakukan transformasi huruf kapital pada data opini menjadi huruf kecil. Hanya menerima 'a' sampai dengan huruf 'z'. Karakter selain huruf dihilangkan atau dianggap sebagai tanda baca (delimiter). Diagram alur mengubah kapitalisasi dapat dilihat pada Gambar 2.



Gambar 2: Diagram alur mengubah Kapitalisasi

Algoritma mengubah kapitalisasi yaitu:

1. Membaca data opini (dataset).
2. Memeriksa ukuran setiap karakter dari awal hingga akhir karakter.
3. Mengubah huruf kapital menjadi huruf kecil.
4. Hasil data opini seluruhnya telah diubah menjadi huruf kecil dan hasil data tahap mengubah kapitalisasi disimpan untuk tahap tokenisasi.

Untuk melakukan tahap mengubah kapitalisasi, berikut penggalan sintaks pada RStudio.
kalimat = tolower(kalimat)

Pada tahap mengubah kapitalisasi, ada beberapa aturan proses agar hasil mengubah kapitalisasi dapat sesuai yang diharapkan. Adapun aturan-aturan tersebut disajikan pada Tabel 1.

Ilustrasi tahap mengubah kapitalisasi yang disajikan pada Tabel 2.

Tabel 1: Tabel Aturan Tahapan Mengubah Kapitalisasi

Kondisi	Aksi
Data opini memiliki huruf kapital [A...Z]	Maka akan mengubah semua huruf kecil [a...z]
Data opini memiliki karakter Simbol	Maka akan Menghapus karakter simbol tersebut dari data komentar
Data opini memiliki huruf kecil	Tidak ada aksi
Data opini memiliki spasi	Tidak ada aksi

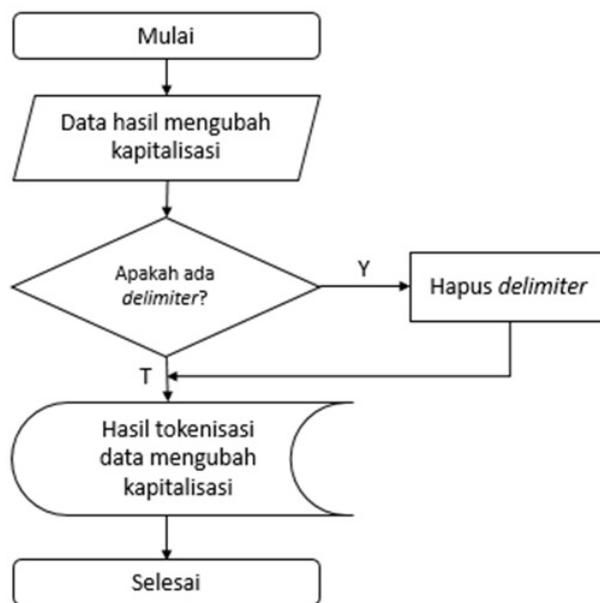
Tabel 2: Tahap Ilustrasi Mengubah Kapitalisasi

Contoh Data			
Data Opini		Hasil Tahapan Mengubah Kapitalisasi	
Input		Output	
Desak Pihak Universitas Gunadarma Memberikan Subsidi Pulsa dan Kuota untuk Mahasiswa		desak pihak universitas gunadarma memberikan subsidi pulsa dan kuota untuk mahasiswa	

Tokenisasi (Tokenizing)

Tokenisasi merupakan proses pemotongan string berdasarkan tiap kata, menghilangkan tanda baca / pembatas (delimiter) seperti

tanda titik (.), tanda koma (,), spasi dan karakter angka yang ada pada kata data opini tersebut. Tahapan ini dilakukan setelah data opini (dataset) melewati tahap mengubah kapitalisasi. Diagram alur tahap (Tokenizing) dapat dilihat pada Gambar 3.



Gambar 3: Diagram Alur Tokenisasi

Algoritma dalam tahap tokenisasi adalah sebagai berikut: 1) Membaca data hasil tahap mengubah kapitalisasi. 2) Memeriksa apakah ada delimiter atau tidak. 3) Menghilangkan tanda baca / pembatas (delimiter) seperti tanda titik (.), tanda koma (,), spasi dan karakter angka. 4) Hasil data tahap mengubah kapitalisasi sudah tidak mengandung delimiter dan hasil data tahap tokenisasi disimpan.

Untuk melakukan tahap tokenisasi, berikut penggalan sintaks pada RStudio.

```
#kalimat = gsub('\\\\w+:(\\\\/\\\\?\\\\/?)[^\\\\s]+', "", kalimat)
```

```
kalimat = gsub('@\\\\w+', "", kalimat)
```

```
kalimat = gsub('#\\\\w+', "", kalimat)
```

```
kalimat = gsub('[:punct:]', "", kalimat)
```

```
kalimat = gsub('[:cntrl:]', "", kalimat)
```

```
kalimat = gsub('\\\\d+', "", kalimat)
```

Pada tahapan tokenisasi (Tokenizing), ada beberapa aturan proses agar hasilnya sesuai dengan yang diharapkan. Aturan- aturan tersebut disajikan pada Tabel 3.

Tabel 3: Tabel Aturan Tahap Tokenisasi

Kondisi	Aksi
Jika data opini yang sudah melalui proses mengubah kapitalisasi mengandung tanda baca / pembatas (delimiter)	Tanda baca / pembatas (delimiter) yang terdapat pada data opini akan dihilangkan
Jika data opini yang sudah melalui proses mengubah kapitalisasi mengandung simbol dan angka	Simbol dan angka yang terdapat pada data opini akan dihilangkan

Kondisi Aksi Jika data opini yang sudah melalui proses mengubah kapitalisasi mengandung tanda baca / pembatas (delimiter) Tanda baca / pembatas (delimiter) yang terdapat pada data opini akan dihilangkan Jika data opini yang sudah melalui proses mengubah kapitalisasi mengandung simbol dan angka Simbol dan angka yang terdapat pada data opini akan dihilangkan

Ilustrasi tahap tokenisasi (Tokenizing) yang disajikan pada Tabel 4.

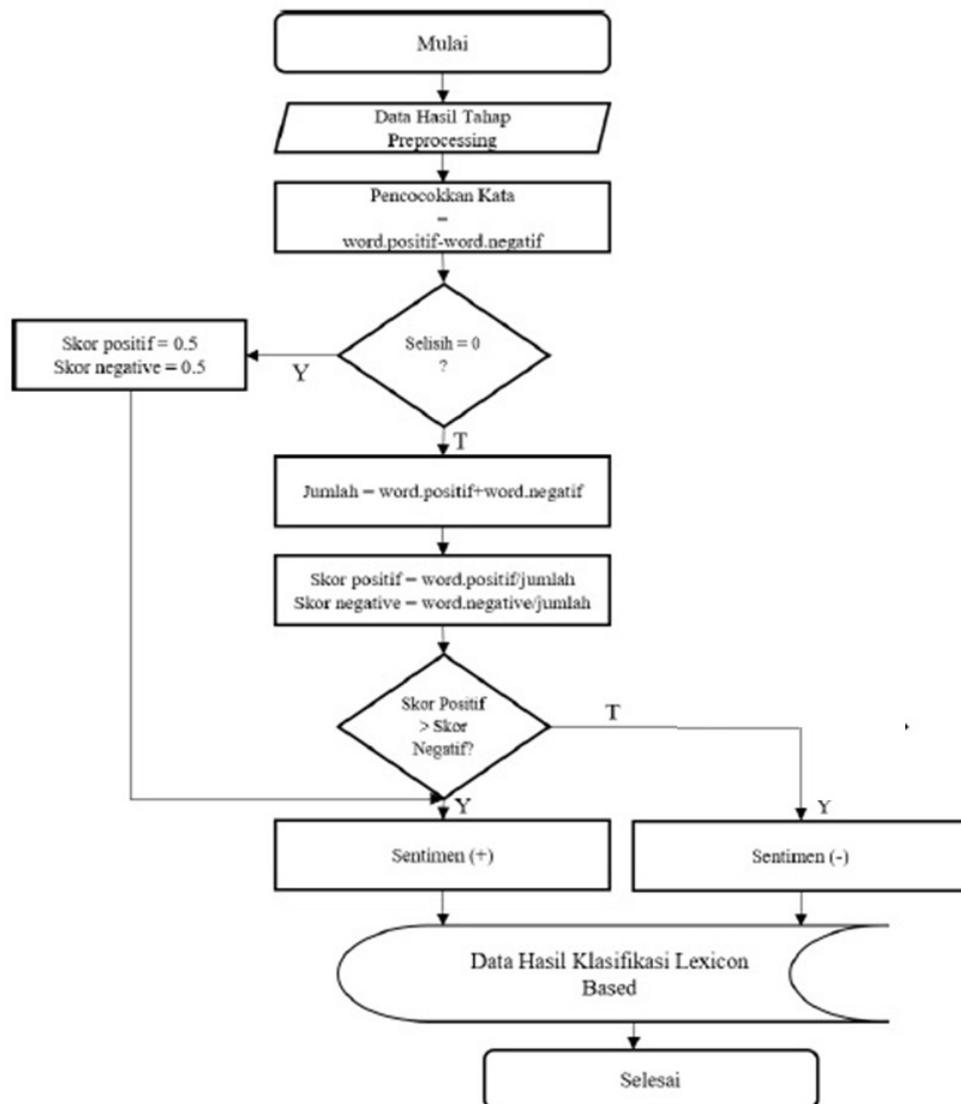
Tabel 4: Tabel Ilustrasi Tahapan Tokenisasi

Contoh Data			
Data Opini Hasil Tahapan Mengubah Kapitalisasi		Hasil Tahapan Tokenisasi	
Input	Universitas gunadarma ciptakan robot ventilator untuk pasien covid-19	Output	Universitas gunadarma ciptakan robot ventilator untuk pasien covid

yang telah melalui tahap pembersihan (pre-processing). Klasifikasi ini mencocokkan kata-kata yang terdapat dalam database kata positif dan negatif pada penelitian sebelumnya. Jika data opini dalam satu kalimat memiliki proporsi kata positif lebih besar dari kata negatif, maka kalimat itu digolongkan pada sentimen positif. Namun jika data opini dalam satu kalimat memiliki proporsi kata negatif lebih besar dari kata positif, maka kalimat itu digolongkan pada sentimen negatif. Diagram alur tahap klasifikasi *Lexicon Based* dapat dilihat pada Gambar 4.

Klasifikasi Lexicon Based

Klasifikasi sentimen dengan lexicon based merupakan klasifikasi berdasarkan kata positif dan kata negatif yang ada dalam kata opini



Gambar 4: Diagram Alur Tahap Klasifikasi Lexicon Based

Algoritma klasifikasi *lexicon based* adalah sebagai berikut:

1. Membaca data hasil tahap preprocessing.
2. Pencocokan kata dengan daftar kata positif dan kata negatif.
3. Dilakukan perhitungan skor.
4. Jika proporsi kata positif dan negatif = 0 (no), data hasil tahap preprocessing termasuk sentimen positif.
5. Jika proporsi kata positif lebih besar daripada kata negatif, maka data hasil tahap preprocessing termasuk sentimen positif.
6. Jika proporsi kata negatif > kata positif, maka data hasil tahap preprocessing termasuk sentimen negatif.
7. Hasil tahap preprocessing yang sudah melalui tahap klasifikasi menggunakan *lexicon based* disimpan untuk tahap klasifikasi *Naïve Bayes Classifier*.

Hasil klasifikasi metode *lexicon based* dapat dilihat pada Gambar 5.

	text	positif	negatif	summary	sentimen
1	good morning guys ð apartemen konsep hotel berada di p...	1	-1	0	positif
2	mana ada kaya ginianððð udh labnya boros bgt yutub mulu	3	-3	0	positif
3	btw pernah kuliah di gunadarma wkwk aku keterima disitu ...	1	-3	-2	negatif
4	maaf banget aku bingung pengen masuk gunadarma mau ...	10	-13	-3	negatif
5	maap ak taunya ui doang ama pancasila dan gunadarma	0	-1	-1	negatif
6	rt upn veteran jakarta vs gunadarma muzakkar phonna vs r...	0	0	0	positif
7	universitas gunadarma margonda terbuka luas	2	0	2	positif
8	itu jawaban dari gambar yang dia kirim dengan mengunaka...	2	-2	0	positif
9	kost dekat ui terima kost putra karyawan mahasiswa di lok...	2	-4	-2	negatif
10	iya mba rektor nya pun mantan rektor gunadarmað	0	0	0	positif

Gambar 5: Hasil Klasifikasi Metode *Lexicon Based*

Database Kata Positif dan Kata Negatif

Penentuan suatu kalimat yang termasuk sentimen positif atau negatif dengan cara menghitung proporsi jumlah kata positif dan negatif dilakukan dengan menggunakan metode *lexicon*

based. Penentuan suatu kata dalam kalimat memiliki makna sentimen positif atau negatif ditentukan berdasarkan kamus / database kalimat positif dan negatif pada penelitian sebelumnya. Berikut ini adalah sebagian kamus / database kata positif dan kata negatif yang dapat dilihat pada gambar 6 dan 7.

Lampiran Sebagian Daftar Kata Positif

1	adil	26	efisien	51	lancar	76	ramah
2	akurat	27	ekonomis	52	layak	77	rapi
3	aman	28	favorit	53	layan	78	realisasi
4	amanah	29	gagah	54	lebih baik	79	responsif
5	anggun	30	gairah	55	lezat	80	sabar
6	asli	31	gembira	56	mahir	81	sahabat
7	asyik	32	gesit	57	maklum	82	saran
8	bagus	33	giat	58	makmur	83	sayang
9	bahagia	34	hadiah	59	malaikat	84	segar
10	baik	35	halal	60	mantap	85	senang
11	bangga	36	halus	61	nikmat	86	setia
12	bantu	37	handal	62	nyaman	87	taat
13	bebas	38	ideal	63	optimal	88	tabah
14	cekatan	39	idola	64	optimis	89	takdir
15	cemerlang	40	imajinatif	65	pantas	90	tampai
16	cepat	41	iman	66	pas	91	unggul
17	cerdas	42	imut	67	pasti	92	unik
18	cerdik	43	jenius	68	patuh	93	untung
19	cocok	44	juara	69	peka	94	utuh
20	damai	45	jujur	70	pelindung	95	variasi
21	dinamis	46	karisma	71	pembela	96	variatif
22	Dukung	47	kasih sayang	72	pembuka	97	visioner
23	edukatif	48	kaya	73	pemulihan	98	wahyu
24	efektif	49	keadilan	74	rahmat	99	waspada
25	efektivitas	50	kebaikan	75	rajin	100	yakin

Gambar 6: Lampiran Sebagian Kata Positif

Lampiran Sebagian Daftar Kata Negatif

1	abnormal	26	egois	51	labil	76	patung
2	absurd	27	ekstremis	52	lalai	77	pecundang
3	acak	28	enggai	53	lama	78	pedih
4	acak-acakan	29	frustasi	54	lamban	79	pembenci
5	acuh	30	gagal	55	lambat	80	racun
6	agresif	31	gagap	56	lancang	81	radikal
7	ancaman	32	galib	57	lemah	82	rapuh
8	bahaya	33	galak	58	mabuk	83	rasis
9	bajingan	34	ganas	59	mafia	84	rendahan
10	bangkrut	35	hambatan	60	mahal	85	rentan
11	basi	36	haram	61	malas	86	represi
12	bau	37	histeris	62	malu	87	represif
13	bejat	38	hutang	63	manipulasi	88	sabotase
14	bekas	39	inflamasi	64	marah	89	sakit
15	cacian	40	iritasi	65	naif	90	salah
16	celaka	41	ironi	66	najis	91	sandera
17	cemas	42	jahat	67	nakal	92	tabu
18	cemberut	43	janggal	68	ngeri	93	tahayul
19	cemburu	44	jatuh	69	obsesif	94	umpatan
20	demam	45	jelek	70	oportunistik	95	usang
21	demoralisasi	46	jengkel	71	oposisi	96	usil
22	dendam	47	kabur	72	otoriter	97	virus
23	dengki	48	kabut	73	paksaan	98	vulgar
24	depresi	49	kacau	74	palsu	99	wabah
25	derita	50	kadaluarsa	75	panik	100	waria

Gambar 7: Lampiran Sebagian Kata Negatif

Klasifikasi *Naïve Bayes Classifier*

Tahap klasifikasi menggunakan metode *Naïve Bayes Classifier* merupakan tahap untuk pengklasifikasian apakah data yang diuji termasuk kedalam sentimen positif atau negatif. Metode ini terdiri dari 2 proses yaitu, proses pelatihan dan proses pengujian. Berikut ini adalah tahapan proses klasifikasi menggunakan metode *Naïve Bayes Classifier*:

1. Mendefinisikan Dataset

Dalam penelitian ini dataset yang digunakan adalah data opini pengguna Twitter sebanyak 267 data yang telah melalui tahap pembersihan (Pre-processing) dan telah dilakukan klasifikasi lexicon based. Dibawah ini pengglan sintaks untuk memanggil file yang menjadi dataset untuk klasifikasi menggunakan metode *Naïve Bayes Classifier*.

```
DatasetGunadarma <- read.csv
("C:\\Users\\ASUS\\Downloads\\sentimen_analisis\\sentimen_analisis\\Gunadarma_Dataset_clean.csv")
```

Untuk menampilkan dataset digunakan

sintaks View dan hasilnya dapat dilihat pada Gambar 8.

Kemudian melihat jumlah dan probabilitas dataset berdasarkan *field* sentimen, lalu menampilkan Panjang karakter pada setiap baris kalimat. Dibawah ini penggalan sintaks untuk melihat jumlah probabilitas *field* sentimen dan menampilkan Panjang karakter pada setiap kalimat pada *field* text.

```
textlength <- str_length(kalimat)
```

Output panjang karakter dengan membuat *field* baru bernama textlength dapat dilihat pada Gambar 9.

2. Pelabelan Data Dalam tahap ini dilakukan proses pelabelan untuk menentukan rentang data yang masuk kategori data latih dan data uji. Proporsi pada pembagian data latih dan data uji sebesar 70:30 (dalam persen) dari 267 data. Field yang digunakan adalah sentimen karena dalam penelitian ini pokok perhitungan sentimen positif atau negatif.

View (DatasetGunadarma)

	text	positif	negatif	summary	sentimen
1	good morning guys ð apartemen konsep hotel berada di p...	1	-1	0	positif
2	mana ada kaya ginianððð udh labnya boros bgt yutub mulu	3	-3	0	positif
3	btw pernah kuliah di gunadarma wkwk aku keterima disitu ...	1	-3	-2	negatif
4	maaf banget aku bingung pengen masuk gunadarma mau ...	10	-13	-3	negatif
5	maap ak taunya ui doang ama pancasila dan gunadarma	0	-1	-1	negatif
6	rt upn veteran jakarta vs gunadarma muzakkar phonna vs r...	0	0	0	positif
7	universitas gunadarma margonda terbuka luas	2	0	2	positif
8	itu jawaban dari gambar yang dia kirim dengan mengunaka...	2	-2	0	positif
9	kost dekat ui terima kost putra karyawan mahasiswa di lok...	2	-4	-2	negatif
10	iya mba rektor nya pun mantan rektor gunadarmað	0	0	0	positif

Gambar 8: Dataset Yang Telah Terdefinisi

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

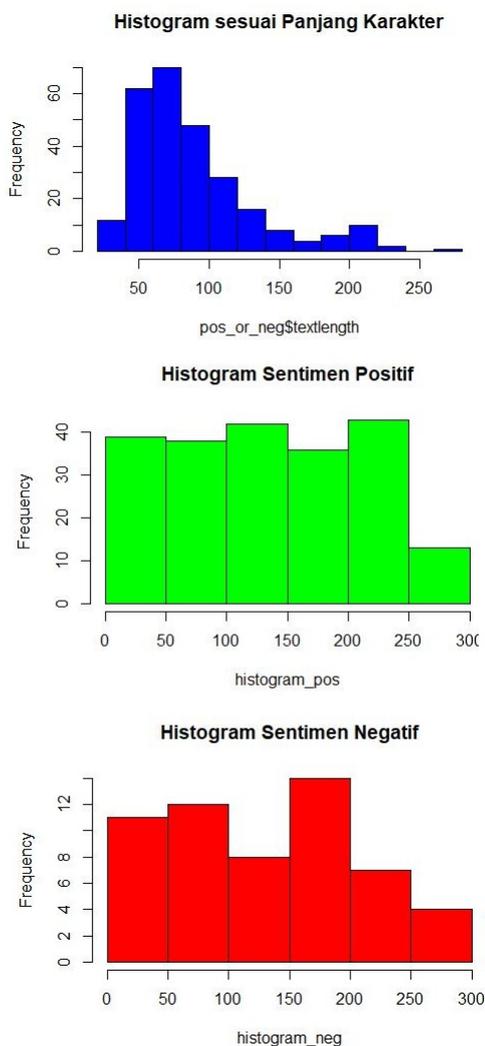
Go to file/function Addins

1.olah_data.R x DatasetGunadarma x DatasetGunadarma.scores x 2.wordcloud.R x 3.pie_chart.R x

Filter

	text	positif	negatif	summary	sentimen	textlength
1	good morning guys ð apartemen konsep hotel berada di p...	1	-1	0	positif	215
2	mana ada kaya ginianððð udh labnya boros bgt yutub mulu	3	-3	0	positif	57
3	btw pernah kuliah di gunadarma wkwk aku diterima disitu ...	1	-3	-2	negatif	85
4	maaf banget aku bingung pengen masuk gunadarma mau ...	10	-13	-3	negatif	265
5	maap ak taunya ui doang ama pancasila dan gunadarma	0	-1	-1	negatif	51
6	rt upn veteran jakarta vs gunadarma muzakkar phonna vs r...	0	0	0	positif	115
7	universitas gunadarma margonda terbuka luas	2	0	2	positif	44
8	itu jawaban dari gambar yang dia kirim dengan mengunaka...	2	-2	0	positif	92
9	kost dekat ui terima kost putra karyawan mahasiswa di lok...	2	-4	-2	negatif	241
10	iya mba rektor nya pun mantan rektor gunadarmað	0	0	0	positif	48

Gambar 9: Field Panjang Karakter



Gambar 10: Histogram Data Sentimen 517

Hasil Visualisasi Data

Visualisasi data digunakan untuk memaparkan hasil dari proses klasifikasi analisis sentimen berdasarkan nilai dalam bentuk histogram, diagram pie, dan wordcloud.

Bentuk Histogram

Data yang dipaparkan merupakan hasil dari setiap tahapan analisis sentimen. Pada Gambar 10 dipaparkan histogram dengan sumbu x sebagai data ke-n dan sumbu y berisi banyak data yang sering muncul, lihat Gambar 10.

Pada Gambar 11 dipaparkan histogram dengan sumbu x sebagai data ke-n dan sumbu y adalah presentase data dari jumlah total data sentimen positif dan negatif untuk parameter panjang karakter.

Pada Gambar 12 dipaparkan histogram dengan sumbu x sebagai data ke-n dan sumbu y adalah presentase data dari jumlah total data sentimen positif atau negatif untuk parameter summary.

Bentuk Diagram Pie

Data yang dipaparkan merupakan proporsi antara sentimen positif dengan negatif dari ke-

Pengujian Akurasi Metode Klasifikasi Menggunakan *Confusion Matrix*

Pengujian akurasi klasifikasi sistem berguna untuk mengetahui tingkat akurasi klasifikasi data tweet dengan menggunakan metode lexicon based dan metode *naïve bayes classifier*. Pengujian ini menggunakan *Confusion matrix* yaitu sebuah matrix prediksi yang dibandingkan dengan kelas dari data inputan asli. Data yang digunakan merupakan data uji yang berjumlah 186 data. Data uji yang sudah melalui tahap klasifikasi menggunakan lexi-

con based dan metode *naïve bayes classifier* dibandingkan. *Crosstable* merupakan sintaks yang tersedia pada RStudio untuk membuat suatu label lalu menghitung probabilitasnya secara vertikal dan horizontal. Tabel merupakan sintaks untuk menampilkan jumlah sentimen data positif dan negatif. Di bawah ini adalah penggalan sintaks pengujian menggunakan *confusion matrix* serta output dari *crosstable* dapat dilihat pada Gambar 16.

Sintaks untuk menghasilkan hasil pengujian dapat dilihat pada Gambar 17.

Output dari akurasi sistem pengujian menggunakan *confusion matrix* dapat dilihat pada Gambar 18.

Cell Contents

			N
	N / Row Total		
	N / Col Total		

Total observations in Table: 81

Actual	Predicted		Row Total
	negatif	positif	
negatif	4 0.800 0.286	1 0.200 0.015	5 0.062
positif	10 0.132 0.714	66 0.868 0.985	76 0.938
Column Total	14 0.173	67 0.827	81

Gambar 16: Hasil Pengujian *Naive Bayes Classifier (Crosstable)*

```
> table("Actual" = pos_or_neg_test_labels, "Predictions" = gunadarma_test_pred)
      Predictions
Actual  negatif positif
negatif    4      10
positif    1      66
```

Gambar 17: Sintaks Pengujian *Naive Bayes*

```

> conf.mat$overall
  Accuracy      Kappa AccuracyLower AccuracyUpper AccuracyNull AccuracyPValue
0.86419753 0.36311651 0.76999037 0.93021131 0.82716049 0.23593361
McNemarPValue
0.01586133

> conf.mat$overall['Accuracy']
Accuracy
0.8641975
    
```

Gambar 18: Hasil Akurasi sistem Menggunakan *Confusion Matrix*

Dari data pengujian akurasi tersebut, diperoleh hasil akurasi klasifikasi dengan menggunakan metode naive bayes classifier sebesar 86.42% dengan kesalahan sistem sebesar 13.58%. Kesimpulan dari pengujian akurasi menggunakan confusion matrix adalah bahwa naive bayes classifier dapat digunakan sebagai metode pengklasifikasian pada penelitian analisis sentimen ini.

Penutup

Penelitian ini telah dilakukan analisis sentimen terhadap Universitas Gunadarma. Proses analisis sentimen dilakukan menggunakan bahasa R dengan IDE RStudio. Kemudian proses klasifikasi dilakukan menggunakan 2 metode yakni: lexicon based untuk menentukan kalimat supaya memiliki makna sentimen positif atau negatif dan naive bayes classifier untuk menghitung akurasi antara prediksi dan hasil aktual dari sistem. Analisis sentimen divisualisasikan dalam bentuk histogram, diagram pie dan wordcloud. Pada penelitian analisis sentimen ini didapatkan hasil akurasi klasifikasi data tweet dari sistem analisis sentimen dengan menggunakan naive bayes classifier sebesar 86.42% dengan kesalahan sistem sebesar 13.58%. Berdasarkan hasil analisis sentimen tersebut dapat disimpulkan bahwa sentimen pengguna Twitter terhadap Universitas Gunadarma memiliki kecenderungan sentimen positif dimana berdasarkan hasil analisis sentimen diperoleh 211 sentimen positif dan 56 sentimen negatif. Hasil analisis ini telah divisualisasikan dalam bentuk histogram, diagram pie dan wordcloud.

Daftar Pustaka

- [1] Puguh Kurniawan, “Pemanfaatan Media Sosial Instagram sebagai Komunikasi Pemasaran Modern Batik Burneh”, Kompetensi, Universitas Trunojoyo, Vol 11, No 2, Oktober 2017, ISSN: 1907-4824 (print), ISSN: 2541-2655(online), pp. 217-225, 2017.
- [2] Paisal, “Analisis Sentimen Masyarakat Berdasarkan Opini dari Sosial Media Menggunakan Metode Naive Bayes Classifier (Study Kasus : Universitas Sjakhyakirti)”, Jurnal Ilmiah Informatika Global, vol. 11 No. 01, Juli 2020, ISSN PRINT : 2302-500X ISSN ONLINE : 2477-3786, pp. 41-46, 2020.
- [3] Nabila Halfi, “Analisis Sentimen terhadap universitas Sriwijaya berdasarkan Opini Masyarakat dari Twitter menggunakan Naive Bayes Classifier”, Program Studi Sistem Informasi, Fakultas Ilmu Komputer, universitas Sriwijaya, 2019.
- [4] Deden Rustiana, Nina Rahayu, “Analisis Sentimen Pasar otomotif Mobil: Tweet Twitter menggunakan Naive Bayes”, Jurnal SIMETRIS, Vol 8 No 1 April 2017, ISSN: 2252-4983, pp. 113-120, 2017.
- [5] Fajar Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter”, Jurnal INOVTEK POLBENG – Seri Informatika, Vol. 3, No. 1 , Juni 2018, ISSN : 2527-9866, 50-59, 2018.
- [6] Bambang Kurniawan, Syahril Effendi, Opim Salim Sitompul, “Klasifikasi Konten Berita Dengan Metode Text Mining”, Journal Dunia Teknologi Informasi, Vol.1 issue 1, 2012.
- [7] Ria Ine Pristiyanti, Mochammad Ali Fauzi, Laili Muflikhah, “Sentimen Analisis Peringkasan Review Film Menggunakan Metode Information Gain dan K -

- Nearest Neighbor”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, ISSN: 2548-964X, Vol. 2 No. 3, Maret 2018, hlm. 1179-1186, Fakultas Ilmu Komputer, Universitas Brawijaya, 2018.
- [8] Jiawei Han, Jian Pei, Micheline Kamber, “Data Mining: Concepts and Techniques”, Third Edition, Morgan Kauffman Publishers, Elsevier, USA, 2011.
- [9] Falahah, Dyar Dwiki Adriadi Nur, “Pengembangan Aplikasi Sentimen Analisis menggunakan Metode Naive Bayes (Studi Kasus Sentimen Analysis dari media Twitter)”, Seminar Nasional Sistem Informasi Indonesia, 2-3 November 2015.
- [10] Lila Dini Utami, Romi Satria Wahono, “Integrasi Metode Information Gain untuk Seleksi Fitur dan Adaboost untuk Mengurangi Bias pada Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes”, *Journal of Intelligent Systems*, Jilid 1, Terbitan 2, Halaman 120-126, 2015.
- [11] Fink, R. Clayton, “Implementasi Opinion Mining (Analisis Sentimen), Coarse- and Fine-Grained Sentimen. Analysis of Social Media Text”, Johns Hopkins APL Technical, 2011.
- [12] Fauziah Afshoh, “Analisa Sentimen menggunakan Naive Bayes untuk Melihat Persepsi Masyarakat terhadap Kenaikan Harga Jual Rokok pada Media Sosial Twitter”, Program Studi Informatika, Fakultas Komunikasi dan informatika, Universitas Muhammdyah Surakarta, 2017.