

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

Tugas Akhir



Dibuat Oleh :

LEO AGUNG PRAMATHANA

140707681

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2019

HALAMAN PENGESAHAN

SENTIMEN ANALISIS PADA TWITTER TERHADAP PEMILIHAN PRESIDEN 2019 MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Yogyakarta, 10 Desember 2019

Leo Agung Pramathana

140707681

Laporan ini telah diperiksa dan disetujui

Oleh :

Dosen Pembimbing 1,



Djoko Budiyanto SHR A. Ir. M.Eng.

KATA PENGANTAR

DAFTAR ISI

HALAMAN PENGESAHAN	2
BAB I.....	10
PENDAHULUAN	10
1.1 Latar Belakang.....	10
1.2 Perumusan Masalah.....	11
1.3 Batasan Masalah	12
1.4 Tujuan Penelitian	12
1.5 Manfaat	13
1.6 Metode Penelitian.....	13
BAB II	14
TINJAUAN PUSTAKA	14
2.1 Tinjauan Pustaka	14
BAB III.....	23
LANDASAN TEORI	23
3.1 Landasan Teori	23
3.1.1 Twitter.....	23
3.1.2 Twitter API	23
1. REST API.....	23
2. Streaming API	24
3. Ads API	24
3.1.3 Data Mining.....	24
3.1.4 Karakteristik Data Mining.....	25
3.1.5 Klasifikasi.....	25
3.1.6 SentiStrength.....	26
3.1.7 WEKA	27
3.1.7 K-Fold Cross Validation.....	27

3.1.8 N-Gram	28
3.1.9 Metode Naïve Bayes Classifier	28
3.1.10 Sentimen Analisis	29
3.1.11 Pengumpulan Data	31
3.1.12 Preprocessing.....	31
3.1.12.1 StopWords	31
3.1.12.2 Steaming	31
3.1.12.3 Case Folding	31
3.1.12.4 Tokenizing	32
BAB IV	33
METODOLOGI PENELITIAN.....	33
4.1. Studi Literatur	34
4.2. Pembentukan Dataset.....	34
a. Data Mining	34
b. Pre-processing Data.....	35
1. Data Cleansing	35
2. Stemming	37
3. Stopword	37
4.3 Pembentukan Model	38
4.3.1 Menentukan <i>Class Attribute</i>	38
4.3.2 Klasifikasi	38
4.4 Evaluasi Model	39
4.4.1 Menentukan Fitur.....	39
4.4.2 Analisa Model.....	39
a. <i>Confusion Matrix</i>	39
b. Akurasi.....	41
c. Presisi.....	41
d. <i>Recall</i>	41
e. F-measure	42
4.5 Penggunaan Cross Validation	43
4.6 Perbandingan Metode Naïve Bayes Classifier dengan Metode Lain	44
4.7 Interpretasi Data	44
4.7.1 Mengidentifikasi Isu pada hasil Pemilihan Presiden 2019	44
4.7.2 Pengelompokan Sentimen dari isu yang berpengaruh terhadap pemilihan presiden 2019.44	
BAB V.....	45
HASIL DAN PEMBAHASAN.....	45

5.1 Dataset.....	45
5.2 Menentukan Class Attribute SentiStrength	48
5.3 Pengujian Dataset	51
5.3.1 Hasil Pengujian dengan metode Naïve Bayes Classifier.....	51
5.3.2 Hasil Pengujian dengan metode lain	53
5.3.3 Perbandingan hasil semua pengujian metode.....	56
5.4 Faktor yang mempengaruhi hasil pilpres 2019	57
5.4.1 isu sara	59
5.4.2 isu radikalisme	60
5.4.3 isu hoax	62
5.4.4 isu kampanye	63
5.5 Analisis sentiment dengan isu yang berpengaruh pada pilpres 2019.....	65
5.5.1 analisis sentiment terhadap isu sara	65
5.5.2 analisis sentiment terhadap isu radikalisme	66
5.5.3 analisis sentiment terhadap isu hoax	68
5.5.4 analisis sentiment terhadap isu kampanye	69
5.6 Analisis hasil dari opini masyarakat	71
<i>BAB VI</i>	72
<i>KESIMPULAN DAN SARAN</i>	72
6.1 Kesimpulan	72
6.2 Saran	72
<i>DAFTAR PUSTAKA</i>	74

Tabel 2. 1 Sample Perbandingan Penilaian Daerah Kemacetan Jakarta	15
Tabel 2. 2 Perbandingan Klarifikasi Positif Dan Negatif dengan metode NBC dan SVM.....	16
Tabel 2. 3 Hasil Pengujian <i>Naïve Bayes Classifier</i>	17
Tabel 3. 1 Contoh hasil SentiStrength	26
Tabel 4.2 Kata kunci pengambilan data Twitter.....	34
Tabel 4.3 Contoh proses data cleansing	36
Tabel 4.4 Contoh Kamus Senti-Strength	38
Tabel 4.5 <i>k-fold cross validation</i>	43

Abstrak

Sentimen Analisis atau bisa disebut *opinion mining* adalah suatu data riset dengan topik yang penting yang saat ini sedang banyak diperbincangkan dari media social. Dan opinion mining bisa meneliti dari data text mining yang dikeluarkan oleh pengguna-pengguna media social Twitter. Dari data tersebut terdapat tiga struktur proses yang dapat di analisis dari *attitude, emotion, opinion*. *Opinion mining* di dalam dunia bisnis banyak sekali digunakan untuk menganalisis opini-opini yang terjadi didalam dunia bisnis tersebut.

Dari penelitian ini akan menganalisis opini-opini pengguna media social twitter pada calon pemilihan presiden 2019. Pada data struktur yang akan di analisis tersebut seperti *attitude, emotion, dan opinion* akan diterapkan peraturan sendiri yang bisa menentukan apakah data struktur tersebut berupa data opini atau bukan, dan menentukan apakah opini tersebut berupa sebagian dari data tekstur tersebut. Data yang sudah dapat akan dianalisis kembali dan di klarifikasikan apakah data tersebut berupa data positif atau negatif menggunakan metode *Naïve Bayes Classifier*.

Kata kunci—Data Struktur, Naïve Bayes Classifier, Sentimen Analisis, Opinion Mining.

Abstract

Sentiment Analysis or can be called opinion mining is a research data with important topics that are currently being discussed from social media. And opinion mining can examine from text mining data released by users of social twitter media. From these data there are three process structures that can be analyzed from attitude, emotion, opinion. Opinion mining in the business world is widely used to analyze opinions that occur in the business world.

From this study will analyze the opinions of social twitter media users in the 2019 presidential selection candidates. In the data structure that will be analyzed such as attitude, emotion, and opinion will be applied by its own rules that can determine whether the structure data is opinion data or not, and determine what is the opinion is part of the texture data. Data that can already be re-analyzed and clarified whether the data is positive or negative data using the Naïve Bayes Classifier method.

Keywords---Data Structure, Naïve Bayes Classifier, Sentiment Analysis, Opinion Mining.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sentimen analisis bisa disebut juga *opinion mining* adalah sebuah proses untuk mengelola *textual data* untuk mendapatkan suatu informasi yang didalamnya terdapat informasi yang *hidden knowledge*. Sentimen analisis tersebut bertujuan untuk menentukan *attitude*, *emotion*, dan opini-opini yang dikeluarkan oleh pengguna media social seperti twitter. Dan twitter sendiri merupakan social media yang fungsinya interaktif untuk pengguna mengkritik semua isu apalagi isu tersebut isu besar seperti pemilihan pilpres 2019 yang memiliki layanan *real time*.

Pemilihan umum presiden 2019 tersebut sangatlah memiliki informasi-informasi data yang sangat banyak dan masih belum beraturan karena semua lapisan masyarakat ikut berkontribusi dan mengeluarkan ekspresi opini mereka dan membagikan di media social seperti twitter. Dan sehingga dari opini yang dibagikan di media social bagi pengguna tersebut akan adanya pro dan kontra yang berupa cacian, saran, dan keluhan menjadi masalah social bagi masyarakat.

Suatu permasalahan social tersebut merupakan kondisi yang tidak diinginkan terjadi di masyarakat sekitar dengan menimbulkan keresahan umum, masalah social ini menurut Rubington dan Weinberg (1989), yaitu sebuah kondisi yang tidak sesuai dengan nilai-nilai yang moral oleh sebagian masyarakat yang setuju dengan melakukan sebuah kegiatan Bersama diperlukan untuk bisa mengubah kondisi itu, dengan kata lain bisa diartikan dengan berusaha saling gotong royong atau saling membantu. Menurut Mills (1959), yaitu suatu masalah social seperti pemilihan presiden digolongkan sebagai keresahan umum, saat masalah itu berpengaruh secara luas dan menjadi bahan perdebatan umum.

Dilihat dari permasalahan yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran maupun keluhan dengan topik pemilihan umum pilpres 2019, sehingga dapat diketahui informasi *sentiment analysis* pengguna sosol media terhadap pilpres 2019. Permasalahan pengklasifikasian data data yang dikeluarkan sentimen ke dalam konten-konten tertentu bisa diselesaikan dengan metode *Naive Bayes*.

Naive Bayes merupakan teknik prediksi berbasis *probabilistik* sederhana dengan asumsi independensi yang kuat pada fitur, dalam artian sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Menurut penelitian yang telah dilakukan sebelumnya, *Naive Bayes* hanya memerlukan sejumlah kecil data untuk mengestimasi *parameter* yang diperlukan untuk menjalankan klarifikasi sentiment analysis tersebut.

Penelitian ini bertujuan untuk mengklarifikasikan permasalahan social tentang pemilihan presiden 2019 pada media social Twitter. Penulis berharap hasil penelitian ini mengklarifikasi permasalahan social tentang pemilihan presiden 2019 dapat teridentifikasi dan berguna bagi masyarakat.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas adalah sebagai berikut :

- a. bagaimana menganalisis opini-opini yang berguna dari pengguna sosial media online tersebut?
- b. Bagaimana tingkat keakuratan dari hasil dataset menggunakan metode naïve bayes classifier
- c. Bagaimana dengan opini masyarakat Indonesia di media social pada tweeter terhadap hasil pemilu presiden 2019

1.3 Batasan Masalah

Batasan-batasan masalah dalam pengerjaan penelitian ini yaitu :

- a. pada penelitian ini, data-data ataupun informasi dalam tweet yang dikeluarkan pengguna ditujukan kepada calon pilpres 2019, dan informasi tersebut mengandung *mention*, *hashtag* dan lain-lain.
- b. penelitian menggunakan metode *naïve bayes* sebagai perbandingan.
- c. penelitian ini hanya bisa menganalisis tweet atau informasi dengan Bahasa Indonesia saja.
- d. tokoh yang diambil dataset nya hanya dibatasi dengan kedua calon pilpres 2019.
- e. Algoritma yang digunakan dalam sentimen analisis ini menggunakan *Naïve Bayes*.
- f. Banyaknya penggunaan kata yang tidak sesuai sehingga menyebabkan variasi Bahasa.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini :

- a. Melakukan pengujian tweet atau opini yang ditujukan untuk tokoh-tokoh yang mencalonkan diri sebagai pilpres 2019.
- b. Menganalisis opini dari pengguna twitter menggunakan sentimen analisis berdasarkan opini, emosi dan sifat.
- c. Mengetahui dan mengklarifikasi terhadap opini saat pemilihan presiden 2019.

1.5 Manfaat

Manfaat dari penelitian ini untuk bisa mengetahui dan mengklarifikasi terhadap opini saat pemilihan presiden 2019 besok di Indonesia.

1.6 Metode Penelitian

Metode dari penelitian ini menggunakan metode Naive Bayes Classifier (NBC) dan metode Support Vector Machine (SVM) yang dimana kedua metode ini akan dibandingkan satu sama lain agar terlihat perbandingannya yang lebih spesifik dengan penilaian suatu opini dari twitter.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Sentimen analisis yaitu sebuah teknik untuk mengetahui opini masyarakat terhadap suatu pembahasan tertentu yang didapat dari sebuah kumpulan data. Contohnya dengan penggunaan media social di masyarakat dengan menggunakan Twitter. Twitter sendiri banyak digunakan oleh masyarakat umum untuk berkomunikasi atau berekspresi menyampaikan opininya. Dengan pertumbuhan teknologi informasi yang semakin canggih berpengaruh pada perubahan cara manusia dalam berkomunikasi terhadap sesamanya. Twitter merupakan media social yang biasa digunakan untuk mencari informasi berupa berita dari siapa saja dan mereka juga bisa mengekspresikan perasaan mereka. Di Twitter sendiri tidak hanya menyampaikan tentang perasaan seseorang saja, tetapi juga dengan mengunggah foto ke media social Twitter dimana *user* sedang berkunjung ke suatu tempat dan menyampaikannya lewat isi tweetnya yang di unggah, dari tweet atau opini yang disampaikan oleh *user* akan menyampaikan suatu sentiment yang akan di analisis apakah tweet atau opini tersebut bersifat positif atau negative. Hal-hal semacam itu merupakan suatu sentimen. Penelitian tentang sentiment analisis sendiri sampai sekarang sudah banyak digunakan. Pada bab ini, penulis akan menjabarkan beberapa penelitian yang sudah dilakukan dengan sentiment analisis tersebut.

Busman, Nurhayati, Nadika Sigit.S, Tifani Shallynda.K (2018) meneliti mengenai Analisa Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java untuk Deteksi Kemacetan Jakarta. Dari penelitian ini mereka mendapatkan kesimpulan dengan Aplikasi Deteksi Kemacetan dapat memberikan informasi kemacetan di Jakarta serta memberikan tingkat kemacetan daerah tersebut. Untuk dapat mengambil Data Twitter, Aplikasi dikoneksikan dengan Flume, Hive, dan Derby dengan membuat *twitter apps* untuk mendapatkan *Consumer Key* (API key), *Consumer*

Secret (API secret), *Access Token*, dan *Access Token Secret*. Kemudian dimasukkan ke dalam konfigurasi *twitter.conf* pada Flume agar Flume dapat melakukan streaming data API Twitter.

Setelah data pada Twitter API dapat di streaming, penulis merancang Aplikasi dengan diagram UML, dan hasilnya fitur-fitur pada Aplikasi Deteksi Kemacetan dapat berfungsi sesuai dengan diagram UML yang dirancang sebelumnya. Pada Aplikasi Deteksi Kemacetan, hasil *streaming* data ditampilkan dan data tersebut dianalisa dengan menggunakan analisa sentiment dan hasilnya yaitu memberikan informasi tingkat kemacetan di Jakarta sesuai daerah yang dicari. Setelah mengetahui tingkat kemacetan suatu daerah, pengujian dilakukan pada Aplikasi Deteksi Kemacetan yang dibandingkan dengan *Google Maps*. Pengujian dilakukan dengan menggunakan 4 sample daerah di Jakarta. Sehingga penulis mendapatkan akurasi sebesar 81%. Akurasi tersebut cukup besar untuk sebuah pengujian perangkat lunak. (Busman, 2018).

Tabel 2. 1 Sample Perbandingan Penilaian Daerah Kemacetan Jakarta.

Daerah	Sample Tweet	Data Google Maps	Akurasi
	Nilai	Nilai	
Semanggi	2.375	2.5	95%
Pancoran	2.467	1.5	61%
Tebet	2	3	67%
Pondok Indah	3	3	100%
Akurasi Rata-rata			81%

Ghulam Asrofi Buntoro (2016) meneliti mengenai Analisis Sentimen Hatespeech pada Twitter dengan Metode *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Dari penelitian ini dengan preprocessing data menggunakan *tokenisasi*, *cleansing* dan *filtering*. Data yang digunakan adalah tweet dalam bahasa

Indonesia dengan tagar *HateSpeech* (*#HateSpeech*), dengan jumlah dataset sebanyak 522 tweet yang didistribusikan secara merata menjadi dua sentimen *HateSpeech* dan *GoodSpeech* menggunakan tokenisasi. Hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi *Support Vector Machine (SVM)* dengan *tokenisasi unigram, stopwords list* Bahasa Indonesia dan *emoticons*, dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai *recall* 66,7% nilai *TP rate* 66,7% dan nilai *TN rate* 75,8%.

Selisih nilai yang sedikit ini karena data yang digunakan seimbang, jadi tidak menyebabkan banyak kesalahan dalam klasifikasi, bila menggunakan data yang tidak seimbang akan menyebabkan *data minority class* yang salah diklasifikasi sebagai *data majority class* pada akhirnya menjadikan selisih nilai menjadi besar. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Support Vector Machine (SVM)* lebih tinggi akurasi untuk klasifikasi sentiment tweet *HateSpeech* Bahasa Indonesia dibandingkan metode klasifikasi *Naïve Bayes Classifier (NBC)*. (Buntoro, 2016)

Tabel 2. 2 Perbandingan Klarifikasi Positif Dan Negatif dengan NBC dan SVM.

Naïve Bayes Classifier (NBC)							
Metode Tokenisasi	Emoticon s	Stopword List	Accurac y (%)	Precisio n (%)	Recall (%)	TP Rate (%)	TN Rate (%)
Unigram	Yes	WEKA	62	62,1	62,1	62,1	60
		Indo	61,8	62,1	61,9	61,9	56,6
	No	WEKA	63	63,1	63	63	60,4
		Indo	62	62,4	62,1	62,1	55,5
Bigram	Yes	WEKA	57,8	59,2	57,9	57,9	78,9
		Indo	57,8	59,2	57,9	57,9	78,9
	No	WEKA	55,3	56,9	55,4	55,4	81,5
		Indo	55,3	56,9	55,4	55,4	81,5
N-gram	Yes	WEKA	62,4	62,7	62,6	62,6	61,9
		Indo	63	63,2	63	63	59,6
	No	WEKA	63,7	63,9	63,8	63,8	61,9
		Indo	63,6	63,7	63,6	63,6	60
Support Vector Machine (SVM)							
Metode Tokenisasi	Emoticon s	Stopword List	Accurac y (%)	Precisio n (%)	Recall (%)	TP Rate (%)	TN Rate (%)
Unigram	Yes	WEKA	61,8	62	61,9	61,9	69,1
		Indo	66,6	67,1	66,7	66,7	75,8
	No	WEKA	63,4	63,5	63,4	63,4	69,8
		Indo	66	66,5	66,1	66,1	75,5
Bigram	Yes	WEKA	55,3	57,5	55,4	55,4	84,5
		Indo	55,3	57,5	55,4	55,4	84,5
	No	WEKA	52,6	53,9	52,7	52,7	85,7
		Indo	52,6	53,9	52,7	52,7	85,7
N-gram	Yes	WEKA	62,4	63,4	62,5	62,5	77,4
		Indo	64,55	66,2	64,6	64,6	81,5
	No	WEKA	63,4	64,3	63,4	63,4	77
		Indo	64,1	65,8	64,2	64,2	81,1

Andre Maureen Pudjajana, Danny Manongga (2018) meneliti mengenai Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan *Naïve Bayes*. Dari penelitian ini hasil sentimen analisis terhadap 500 data uji menunjukkan bahwa nilai sentimen negatif sangat tinggi yaitu 68.4%. Sedangkan hasil perbandingan akurasi kedua metode adalah metode *Naïve Bayes* sebesar 87.48% dan *K-Nearest Neighbor* 85.40% dimana metode *Naïve Bayes* lebih akurasi dibanding metode *K-Nearest Neighbor*. Hasil sentimen tersebut dapat diartikan bahwa Twitter menjadi salah satu media berkomunikasi oleh kaum homoseksual di Indonesia. Kehadiran Twitter membantu kaum homoseksual tersebut untuk membuka diri dan saling berbagi hal-hal negatif dengan sesama kaum tersebut. Penyebaran informasi, foto, dan video yang berkaitan dengan pornografi termasuk pornografi homoseksual dapat ditemukan dengan mudah di Twitter. Peningkatan tersebut tentunya akan memicu terjadinya kejahatan seksual dan berbagai penyimpangan lain. Hasil penelitian ini mengingatkan kepada para pengguna Twitter untuk lebih selektif dan kritis dalam penggunaannya. (Andre Maureen Pudjajana, 2018)

Tabel 2.3 Hasil Pengujian *Naïve Bayes Classifier*.

<i>Hashtag</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-score</i>
#gayindonesia	90.81%	87.5%	84.85%	86.15%
#rtgay	84.85%	77.78%	70.00%	73.68%
#gaybrondong	90.00%	95.45%	70.00%	80.77%
#gaybocah	83.84%	89.47%	54.83%	68.00%
#pijatgay	87.88%	92.31%	70.59%	80.00%
Total	87.48%	88.50%	70.05%	77.72%

Table 2.4 Hasil Pengujian *K-Nearest Neighbor*.

<i>Hashtag</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-score</i>
#gayindonesia	91.00%	72.27%	100.00%	84.21%
#rtgay	85.00%	51.85%	87.50%	62.12%
#gaybrondong	81.00%	13.63%	100.00%	24.00%
#gaybocah	85.00%	36.84%	70.00%	48.28%
#pijatgay	85.00%	46.15%	92.31%	61.53%
Total	85.40%	44.15%	89.96%	56.03%

Ahmad Fathan Hidayatullah, Azhari SN (2014) meneliti mengenai Analisis Sentimen Dan Klarifikasi Kategori Terhadap Tokoh Publik Pada Twitter. Dari penelitian ini telah berhasil membangun model untuk melakukan klasifikasi tweet atau opini berdasarkan sentimen dan kategori dengan *Naive Bayes Classifier*. Hasil akurasi pengujian klasifikasi dengan fitur *term frequency* diperoleh sebesar 70,88% sedangkan fitur *TF-IDF* didapatkan akurasi sebesar 69,30%. Klasifikasi menggunakan *tools RapidMiner* dengan *Naive Bayes Classifier* dan fitur *term frequency* diperoleh sebesar 73,81% sedangkan dengan fitur *TF-IDF* diperoleh sebesar 71.11%. Klasifikasi dengan *Support Vector Machine* menghasilkan akurasi 83,14% untuk fitur *term frequency* dan 82,69% untuk fitur *TF-IDF*. Hasil pengujian pada aplikasi yang dibangun dan pada *tools RapidMiner* memperlihatkan bahwa akurasi dengan fitur *term frequency* memberikan hasil akurasi yang lebih baik daripada akurasi dengan fitur *TF-IDF*. Metode *Support Vector Machine* menghasilkan akurasi performansi yang lebih baik daripada metode *Naive Bayes Classifier*. Penggunaan metode *Support Vector Machine* dan *Naive Bayes Classifier* sama-sama memiliki hasil akurasi yang cukup baik untuk klasifikasi hasil tweet. (Ahmad Fathan Hidayatullah, 2014)

Tabel 2.5 Perbandingan Hasil Perhitungan Akurasi.

Metode	Fitur	Akurasi pada Aplikasi	Akurasi pada RapidMiner
<i>Naive Bayes</i>	<i>Term frequency</i>	70,88%	73,81%
	TF-IDF	69,30%	71.11%
SVM	<i>Term frequency</i>	-	83,14%
	TF-IDF	-	82,69%

Tabel 2.6 Pembandingan Penelitian

No	Peneliti	Tahun	Judul	Metode	Hasil
1	Busman, Nurhayati, Nadika Sigit.S, Tifani Shallynda.K	2018	Analisa Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java untuk Deteksi Kemacetan Jakarta	Menghubungkan aplikasi Flume, Hive, dan Derby. Untuk mendapatkan <i>Consumer Key</i> (API key), <i>Consumer Secret</i> (API secret), <i>Access Token</i> , <i>dan Access</i> <i>Token Secret</i>	penulis mendapatkan akurasi sebesar 81%. Akurasi tersebut cukup besar untuk sebuah pengujian perangkat lunak di 4 sampel daerah Jakarta
2	Ghulam Asrofi Buntoro	2016	Analisis Sentimen Hatespeech pada Twitter dengan Metode <i>Naïve</i> <i>Bayes</i> <i>Classifier</i> (NBC) dan <i>Support Vector</i> <i>Machine</i> (SVM)	Menggunakan metode <i>Naïve</i> <i>Bayes Classifier</i> (NBC) dan <i>Support Vector</i> <i>Machine (SVM)</i> dengan <i>tokenisasi</i> <i>unigram</i> , <i>stopword list</i> Bahasa Indonesia dan <i>emoticons</i>	Hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi <i>Support Vector</i> <i>Machine</i> (SVM) dengan <i>tokenisasi</i> <i>unigram</i> , <i>stopword list</i>

					<p>Bahasa Indonesia dan <i>emoticons</i>, dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai recall 66,7% nilai TP rate 66,7% dan nilai TN rate 75,8%</p>
3	Andre Maureen Pudjajana, Danny Manongga	2018	<p>Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan <i>Naïve Bayes</i></p>	<p>Menggunakan metode <i>Naïve Bayes Classifier</i> dan <i>K-Nearest Neighbor</i></p>	<p>nilai sentimen negatif sangat tinggi yaitu 68.4%. Sedangkan hasil perbandingan akurasi kedua metode adalah metode <i>Naïve Bayes</i> sebesar 87.48% dan <i>K-Nearest</i></p>

					<i>Neighbor</i> 85.40% dimana metode <i>Naïve</i> <i>Bayes</i> lebih akurasi dibanding metode <i>K-Nearest Neighbor</i>
4	Ahmad Fathan Hidayatullah, Azhari SN meneliti mengenai	2014	Analisis Sentimen Dan Klarifikasi Kategori Terhadap Tokoh Publik Pada Twitter	Menggunakan Metode <i>Naive Bayes Classifier</i> , <i>TF-IDF</i> dan menggunakan <i>Tools RapidMiner</i> dengan fitur <i>term frequency</i>	Hasil akurasi pengujian klasifikasi dengan fitur <i>term frequency</i> diperoleh sebesar 70,88% sedangkan fitur <i>TF-IDF</i> didapatkan akurasi sebesar 69,30%. Klasifikasi menggunakan <i>tools RapidMiner</i> dengan <i>Naive</i>

					<i>Bayes Classifier</i> dan fitur <i>term frequency</i> diperoleh sebesar 73,81% sedangkan dengan fitur <i>TF-IDF</i> diperoleh sebesar 71.11%
--	--	--	--	--	--

BAB III

LANDASAN TEORI

3.1 Landasan Teori

3.1.1 Twitter

Twitter merupakan media social *Microblogging* dimana pengguna akan berinteraksi secara *realtime* dan membagikan opini yang pengguna buat ke media social. Twitter sendiri digunakan untuk mencari informasi berupa berita terbaru, dan menjadi tempat untuk mengungkapkan suatu perasaan penggunanya. Disebut *Microblogging* dikarenakan pengguna dapat mengirimkan sebuah opini dan membaca pesan layaknya *blog* pada umumnya. Twitter sendiri memiliki format penulisan karakteristik unik dengan menggunakan simbol ataupun aturan khusus lainnya. Di media social Twitter tidak hanya bisa mengungkapkan perasaan pengguna saja, tetapi pengguna juga bisa mengupload foto dan video untuk memperlihatkan lokasi mereka pada saat ini dan juga pengguna juga bisa menambah *hashtag* sendiri di dalam opini yang mereka ungkapkan untuk mengelompokkan konten yang mereka buat dan untuk mempermudah pencarian dari tweet yang mereka ungkapkan. Dari tweet yang pengguna buat itu akan menghasilkan suatu sentimen, dan suatu sentimen ini ketika dianalisis akan mendapatkan sentimen yang bersifat positif atau negatif.

3.1.2 Twitter API

1. REST API

Fungsi *REST API* yaitu menyediakan pengaksesan program Twitter untuk membaca dan menulis opini di Twitter. *REST API* bisa mengidentifikasi media social Twitter menggunakan tanggapan yang tersedia di *JSON*.

2. Streaming API

Streaming API yaitu sebuah implementasi dari pengguna *Streaming* akan meneruskan pesan yang menunjukkan opini-opini lain sudah berjalan dengan pengambilan *endpoint REST*.

3. Ads API

Ads API yaitu sebuah platform iklan yang ada di Twitter bisa mengintegrasikan sebuah iklan yang akan muncul di *recent* Twitter agar tidak mengganggu pengguna dengan pemakaian media social Twitter perlu dikelola oleh *Mitra* Twitter.

3.1.3 Data Mining

Data mining yaitu sebuah proses untuk mendapatkan atau menggali beberapa informasi yang ingin dicari berupa data secara manual dari suatu database, tujuan dari data mining tersebut bertujuan untuk mengambil data yang diperlukan menjadi sebuah informasi yang bisa dibuat lebih berharga, data tersebut dapat cari dengan cara mengekstrak dan menarik data dari dalam database.

Data mining juga bisa disebut dengan nama lain seperti Knowledge discovery in database yang bisa disingkat dengan (KDD), knowledge extraction, business intelligence, dari data mining tersebut merupakan data yang dapat di manipulasi untuk membuat informasi yang lebih penting dengan kebutuhan pemakai yang bertujuan untuk membantu sebuah pengamatan perilaku dari media social seperti twitter sendiri yang kebanyakan orang meluapkan sebuah opini disana dengan menggunakan tweet yang ditulis sendiri oleh kebanyakan orang. Dari definisi data mining tersebut dapat dijabarkan sebagai berikut :

- a. sebuah proses pengambilan data yang besar dari sebuah database yang tersimpan.
- b. Memproses semua data yang diambil tadi dengan jumlah yang besar
Dan proses data mining tersebut sangatlah diperlukan untuk mengolah data yang sangat banyak, dari data – data tersebut yang sudah di mining dari twetter akan

3.1.4 Karakteristik Data Mining

Secara teknis juga data mining disebut sebuah pemanfaatan data – data teknik statistic, matematika, dan kecerdasan buatan untuk mengidentifikasi informasi, berbagai macam pola dari teknik penambangan data itu bisa dimanfaatkan untuk aturan bisnis, kebanyakan definisi data mining termasuk dengan proses yang susah untuk dipahami untuk mengidentifikasi data tweet yang menjadi suatu informasi yang dapat dipahami dan dapat diperjelas dari data – data yang mentah menjadi informasi yang sangat berharga. Berikut karakteristik dan tujuan data mining :

- a. *Supervised*, adalah variabel atau data yang berlabel.
- b. *Semi Supervised*, adalah variabel atau data yang beberapa berlabel dan beberapa tidak berlabel.
- c. *Unsupervised*, adalah variabel yang tidak berlabel.

3.1.5 Klasifikasi

Klasifikasi sendiri adalah sebuah proses menemukan fungsi untuk menjelaskan konsep dan membedakan suatu objek data yang bertujuan untuk mempermudah pengelompokan data sendiri dari suatu kelas yang berbeda agar bisa diberi label sesuai dengan kebutuhan klarifikasi data yang diinginkan, klarifikasi data sendiri terdapat dua proses, yaitu :

- a. Proses training

Proses training bertujuan untuk membangun model dari dataset untuk memberi label label dari sebuah data yang sudah di ambil

b. Proses testing

Proses testing bertujuan untuk memperkuat keakuratan model yang sudah diproses pada saat proses training tadi.

3.1.6 SentiStrength

Sentistrength yaitu prosos klarifikasi algoritma yang menggunakan kamus lexicon dengan sebuah aturan dan informasi tambahan untuk bisa mendeteksi keakuratan sentiment dalam sebuah data text dalam Bahasa Inggris dan Bahasa Indonesia. SentiStrength sendiri bisa mengidentifikasi sebuah agrumen yang bersifat nilai positif dan negatif dari sebuah kata (dual-scale), SentiStrength menggunakan system dual-scale yang bertujuan untuk merasakan sebuah emosional positif dan negatif yang dikeluarkan dari text opini dari sebuah topik yang sudah ditentukan, dan dari jangkauan SentiStrength sendiri ditentukan nilai dari angka 1 sampai angka 5, yang berarti nilai dari angka 1 tersebut menunjukkan bahwa kalimat tersebut tidak memiliki sentiment positif ataupun negatif, dan sedangkan nilai 5 dari kalimat tersebut menunjukkan bahwa kalimat tersebut mempunyai sentimen yang sangat kuat dan mempunyai nilai tertinggi dari positif maupun negatif.

Tabel 3. 1 Contoh hasil SentiStrength

No	Status
1	prabowo kalah [-5] akibat tampang boyolali wkwk
2	aduh [-4] sakit [-3] kalah [-5]
3	sukses [4] prabowo moga jd presiden yg baik [4]

Dari table tersebut angka yang ada di dalam tanda kurung adalah nilai penunjuk kekuatan sentiment dari setiap kata yang sudah sesuai dengan kamus sentiment.

Table 3.2 Peraturan SentiStrength

No	Status
1	Nilai Positif > Nilai Negatif = Positif
2	Nilai Positif < Nilai Negatif = Negatif
3	Nilai Positif = Nilai Negatif = Netral

Tidak hanya kamus sentimen dengan pembobotan oleh manusia yang digunakan, SentiStrength juga bisa menggunakan kamus emotikon dan kamus ungkapan yang telah diberi bobot oleh manusia. (Wahid & SN, 2016)

3.1.7 WEKA

Arti dari WEKA adalah (Waikato Enviroment for Knowladge Analysis). WEKA sendiri adalah sebuah perangkat lunak penyedia implementasi algoritma pembelajaran yang dapat diterapkan dengan mudah pada dataset. Dari perangkat lunak WEKA sendiri memiliki beberapa tools untuk masalah pengelolaan data, mulai dari *regression*, *pre-processing*, *classification*, *clustering*, *association rules*, dan *visualization*. WEKA merupakan perangkat lunak yang dibangun menggunakan Bahasa pemrograman Java yang kemudian didistribusikan menjadi perangkat lunak *open source*(Frank, Hall, & Witten, 2016).

3.1.7 K-Fold Cross Validation

K-Fold Cross Validation adalah suatu teknik yang membuat sebuah validasi pada dataset untuk menentukan akurasi dari sebuah proses pembelajaran. Yang terkenal dengan K-Fold Cross Validation sendiri adalah suatu metode yang bisa

melipat ganda kan data sebanyak yang kita perlukan (K) dan akan mengulangi sebuah proses sebanyak K yang diperlukan tadi. Dari semua nilai total yang dibutuhkan ini akan dicari nilai rata-ratanya dengan rata-rata tertinggi yang nanti nya akan dipilih sebagai model. Metode *k-fold cross validation* juga memiliki kelebihan dan kekurangan sendiri, kelebihan nya dari metode ini yaitu dapat mengklasifikasi *dataset* lebih akurat dan lebih efisien, namun kelemahan nya dari metode ini adalah saat melakukan proses komputasi harus menggunakan data yang lebih besar, karena sebuah proses menggandakan data agar lebih akurat membutuhkan proses lebih banyak sebanyak k kali (Haltuf, 2011).

3.1.8 N-Gram

N-Gram yaitu salah satu proses yang digunakan dalam text mining atau pengolahan kata. Secara umum N-Gram digunakan dalam berbagai macam proses untuk pengolahan bahasa Indonesia untuk mengurutkan sebuah kata yang sering muncul pada sebuah data text

3.1.9 Metode Naïve Bayes Classifier

Dalam sentiment analisis untuk menganalisis sebuah topik pembicaraan yang ditentukan menggunakan berbagai macam metode seperti contoh metode *Naïve Bayes Classifier*, metode ini merupakan suatu teknik untuk memprediksi sebuah hal *probabilistik* yang berdasarkan pada penerapan *theorema bayes* yang diperkuat oleh fitur-fitur yang di asumsikan oleh *theorema bayes*, dengan arti dari fitur-fitur yang diasumsikan oleh *theorema bayes* dari data tersebut tidak berkaitan dengan nada atau tidaknya fitur-fitur lain dalam data yang sama. Metode *Naïve Bayes Classifier* sendiri mempunyai kemampuan untuk klarifikasi data serupa dengan *Decision Tree* dan *Neural Network*. Dan metode

Naïve Bayes Classifier telah terbukti mempunyai akurasi kecepatan tinggi yang dapat diaplikasikan kedalam *database* dengan data yang besar.

Metode Naïve Bayes merupakan klasifikasi metode yang berkelanjutan pada Teorema Bayes. Metode pengklasifikasian ini menggunakan dari sebuah metode probabilitas dan statistik yang ditemukan oleh ilmuwan inggris yaitu, Thomas Bayes. Dari ciri-ciri Naïve Bayes Classifier yaitu mengasumsikan yang lebih kuat akan independensi dari masing-masing kondisi atau kejadian. Naïve Bayes sangat berguna untuk pembelajaran algoritma, pengetahuan masa lalu dan pengamatan data yang bisa digabungkan. Naïve Bayes Classifier menggunakan data dari data training yang digunakan untuk membentuk sebuah model classifier, representasi pengetahuan yang akan digunakan untuk prediksi data baru yang belum ada. Setelah itu membutuhkan data testing yang digunakan untuk mengukur keberhasilan classifier dan klasifikasi dengan benar. Berikut rumus dari Naïve Bayes yang digunakan, yaitu :

$$P(H_i|X) = \frac{P(X|H_i)P(H_i)}{P(X)}$$

Keterangan : X : Kriteria suatu kasus berdasarkan masukan.

H_i: Kelas solusi pola ke-i, dimana i adalah jumlah label kelas.

P(H_i|X) : Probabilitas kemunculan label kelas H_i dengan kriteria masukan X.

P(X|H_i) : Probabilitas kriteria masukan X dengan label kelas H_i.

P(H_i) : Probabilitas label kelas H_i.

3.1.10 Sentimen Analisis

Sentimen analisis atau bisa disebut dengan opinion mining yang memiliki suatu proses untuk memahami dan mengklarifikasi dan mengelolah data data tersebut

kedalam bentuk textual data dengan tujuan agar gampang memperoleh informasi dari data data tersebut. Sentimen analisis sendiri bisa disebut cara untuk mengekstraksi pendapat *opinion*, *attitude*, dan *emotion* orang mengenai suatu topik pembahasan tertentu. Dari menganalisis suatu topik pembahasan tertentu tersebut ada dua sifat yang berlawanan yaitu bersifat *positif* atau bersifat *negatif*.

Menurut B.Pang dan L.Lee sentiment analisis terbagi menjadi 2 kategori, yaitu :

1. *Coarse-grained sentiment analysis*
2. *Fined-grained sentiment analysis*

Pada *Coarse-grained sentiment analysis* sebuah proses klasifikasi dilakukan berdasarkan orientasi sebuah dokumen secara keseluruhan. Di dalam orientasi tersebut dibagi menjadi 3 jenis yaitu negatif, netral, dan positif.

Dan untuk *Fined-grained sentiment analysis* berbeda dengan *Coarse-grained sentiment analysis* dikarenakan metode ini menggunakan objek yang berupa sebuah kalimat melainkan bukan sebuah dokumen secara keseluruhan.

Contoh :

1. Saya senang menjadi bagian dari Atma Jaya Yogyakarta (positif)
2. Teman saya kecewa dengan pelayanan operator Telkomsel (Negatif)

Dari masalah yang sering terjadi di dalam sentiment analysis ini adalah banyaknya penggunaan kata-kata yang tidak sesuai sehingga menyebabkan variasi Bahasa Indonesia. Sebagai contoh kata “kamu” yang menyebabkan variasi bahasa seperti “km”, “kau”, “kmu”, “u”, “lu”, “lo” dan sebagainya.

3.1.11 Pengumpulan Data

Tahap ini bertujuan untuk mencari dan mengumpulkan data. Di penelitian ini pengambilan suatu data menggunakan metode *crawling* dari media social Twitter. Saat data tersebut sudah didapat, akan lanjut ke tahap *Preprocessing*.

3.1.12 Preprocessing

Preprocessing adalah dimana proses untuk mempersiapkan data dan dibersihkan dari dataset untuk selanjutnya di klarifikasi lagi. Dari proses ini terdiri dari subproses, subproses tersebut berupa :

3.1.12.1 StopWords

Proses *StopWords* ini berfungsi untuk menghilangkan kata-kata yang tidak berpengaruh apapun dari *ekstraksi* sentiment suatu tweet, sebagai contoh kalimat yang dimaksud seperti (Yonathan dan exan, 1996) cari contoh kalimat tersebut harus di removal terlebih dahulu karena tidak berkontribusi banyak pada penulis, dengan contoh kata “di”, “dan”, “dengan”, ”ke” dan lain-lain.

3.1.12.2 Steaming

Proses *Steaming* ini berfungsi untuk mengubah imbuhan kata menjadi kata baku. Sebagai contoh kata yang memiliki imbuhan seperti “melambatkan” menjadi “lambai”.

Proses *Steaming* ini berfungsi memudahkan proses melabelkan kata menjadi positif atau negatif dengan bantuan kamus *lexicon*.

3.1.12.3 Case Folding

Proses *Case Folding* sendiri berfungsi menghilangkan angka dan bentuk tanda baca sehingga data yang diambil hanya mengandung karakter huruf a sampai z.

3.1.12.4 Tokenizing

Proses *Tokenizing* ini berfungsi untuk pemisah kata yang ada di dalam kalimat, *Tokenizing* ini tergantung pada karakter spasi yang ada di dalam kalimat untuk proses pemisahan kata.

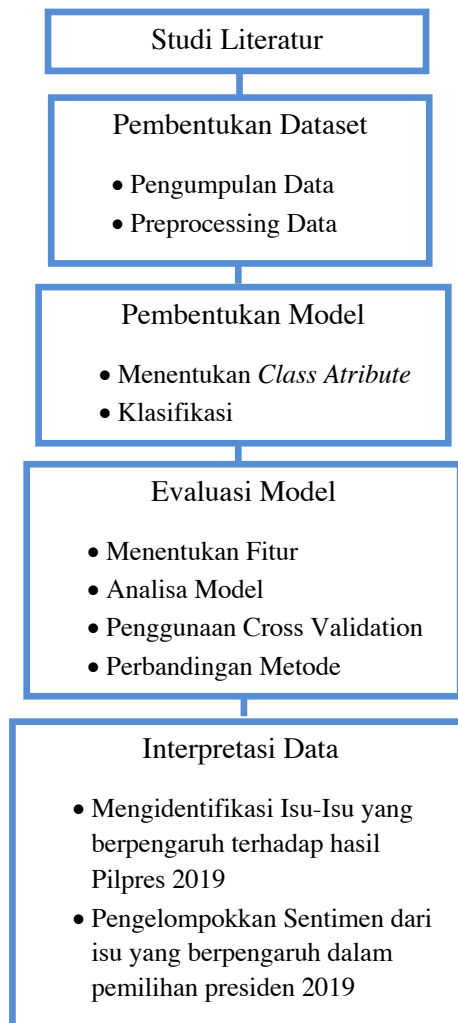
Sebagai contoh “aku tidak ingin golput untuk tahun 2019 ini” dari situ kita bisa melihat ada 8 kalimat yang harus dipisah pisah dengan spasi “aku”, “tidak”, “ingin”, “golput”, “untuk”, “tahun”, “2019”, “ini”.

Itulah, yang menjadi pacuan untuk pemisah antar kalimat adalah spasi dan tanda baca.

BAB IV

METODOLOGI PENELITIAN

Bab empat ini berisikan sebuah pembahasan mengenai metodologi pada penelitian. Tahap pertama adalah kita harus mengetahui studi literature nya. Setelah itu melanjutkan pembentukan *dataset*, setelah itu melanjutkan pembentukan model. Dan tahapan terakhir adalah mengevaluasi sebuah model. Dapat dilihat pada gambar 4.1 mengenai struktur metodologi dalam penelitian ini.



Gambar 4.1 gambar Metodologi Penelitian

4.1. Studi Literatur

Studi Literatur adalah mencari metode dan belajar dari sumber referensi yang bersangkutan dengan penelitian ini, dalam penelitian ini akan meliputi *Data Mining*, Analisis Sentimen, *Naïve Bayes Classifier*, dan dari data *Twitter* yang berhubungan dengan hasil Pemilihan Presiden 2019 kemarin.

4.2. Pembentukan Dataset

Untuk pembentukan dataset, ada beberapa proses yang harus diperhatikan untuk mendapatkan hasil yang lebih akurat. Untuk proses pertama yaitu *data mining*. Proses berikutnya yaitu *preprocessing*, pada proses *preprocessing* ini akan mengubah suatu bentuk data dari data mentah yang belum terstruktur menjadi data yang rapi dan terstruktur. Dan untuk proses yang terakhir, pemberian label positif, netral, dan negative untuk sebuah data text.

a. Data Mining

Pada penelitian ini pengumpulan data sendiri menggunakan aplikasi dari media sosial Twitter, dari pengumpulan data tersebut menggunakan berbagai kata kunci, sebagai contoh kata kunci nya adalah, hashtag yang populer di twitter dengan pembahasan yang sama, waktu awal pengambilan data, dan waktu akhir pengambilan data.

Kata Kunci	"#2019gantipresiden", "#jokowisekalilagi", "#17apriljokowipresiden", "#jokowi2periode", "#prabowopresidenku"
Batasan waktu	01 Agustus 2018 – 31 April 2018
Bentuk data	ID, username, status, tanggal, retweet, favorites, mention, hastags.

Tabel 4.2 Kata kunci pengambilan data Twitter

Total data yang berhasil dikumpulkan sebesar 303.019 ribu *tweet*. Dan untuk contoh hasil dari pengumpulan data adalah sebagai berikut:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	date	username	to	replies	retweets	favorites	text	geo	mentions	hashtags	id	permalink				
2	30/12/18 23.59	bro5871	fadizon	0	0	0	0 Slap, kita siap menuju smngat baru di tahun baru #2019GantiPri @maspyauu @Anggraini #2019GantiPresiden				1.08E+23	https://twitter.com/bro5871/status/1079527531353493504				
3	30/12/18 23.59	randaniy7	durimulya	0	0	0	No comment! Waras geser sebelah... #2019GantiPresiden #2019PrabowoPresidenhttps://tww #2019GantiPresiden				1.08E+23	https://twitter.com/randaniy7/status/1079527499728482305				
4	30/12/18 23.59	heidar171		0	0	1	Coba bong, cebong aye sini ngumpul... Bener gak sih????????? #2019GantiPresiden pic.twitter.com/heidar171/status/1079527455169550336				1.08E+23	https://twitter.com/heidar171/status/1079527455169550336				
5	30/12/18 23.59	Lumpia	panca66	0	0	0	Eesa #2019GantiPresidenhttps://twitter.com/panca66/status/107950799430981856_A# #2019GantiPresiden				1.08E+23	https://twitter.com/Lumpia/status/107952748131391114				
6	30/12/18 23.59	meifa80	sanduno	0	0	0	#2019GantiPresiden #2019Prabowadandi IndonesiaKdHakamur #TahunBaruPresidenBaru #2019GantiPresiden				1.08E+23	https://twitter.com/meifa80/status/1079527395537823792				
7	30/12/18 23.58	randaniy7	ekajumilar	0	0	0	0 Semoga niat dari para calon pemimpin bangsa ini tulus lillahi'ta'ala... Pemerintahlah bersih #2019GantiPresiden				1.08E+23	https://twitter.com/randaniy7/status/107952729455145984				
8	30/12/18 23.58	SoeghartoAque	choetowental	0	2	3	Berarti yg menyuruhnya itu lebih gila daripada orang gila yg disuruhnya itu Makanya kita har #2019GantiPresiden				1.08E+23	https://twitter.com/SoeghartoAque/status/10795271883476				
9	30/12/18 23.58	Lumpia	Awang_shiliang	0	0	0	#2019GantiPresidenhttps://twitter.com/Awang_shiliang/status/1079222313128339456_A# #2019GantiPresiden				1.08E+23	https://twitter.com/Lumpia/status/1079527152603717632				
10	30/12/18 23.58	medaribabun	medaribabun	0	0	0	#2019PrabowoPresidenKita #2019PrabowadandiKita #Lumpia Kita @#n# #Mekah @gustafdev #2019PrabowPresi				1.08E+23	https://twitter.com/medaribabun/status/1079527098855995				
11	30/12/18 23.58	MarcellanaJ	BangPinu	0	0	0	#2019GantiPresiden kali mestinya tunjakin ajah itu bpk nganu diacara debat...berkualitas gak #2019GantiPresiden				1.08E+23	https://twitter.com/MarcellanaJ/status/10795270706474470				
12	30/12/18 23.57	heidar171	IreneViana	0	0	0	0 Coba diocokin #2019GantiPresiden pic.twitter.com/Bxfiqdc6				1.08E+23	https://twitter.com/heidar171/status/107952699065699584				
13	30/12/18 23.57	Lumpia		0	0	1	#2019GantiPresiden				1.08E+23	https://twitter.com/Lumpia/status/1079526935607160025				
14	30/12/18 23.57	Lumpia	MichelanlaAdam	0	0	0	#2019GantiPresidenhttps://twitter.com/MichelanlaAdam/status/107947654212253344_A# #2019GantiPresiden				1.08E+23	https://twitter.com/Lumpia/status/1079526893580279808				
15	30/12/18 23.57	Pono212		0	0	0	Sad Dida: Saya Takla Merencanakan Mobil Lemka, Karena Saya Yakin Itu Kebahagiaan ya #2019GantiPresiden				1.08E+23	https://twitter.com/Pono212/status/10795268400210752000				
16	30/12/18 23.56	TgVip		0	0	0	Prediksi tagel Singapore Senin 31 Desember 2018 Kepala : 6.4.7.3 Angka Ikat : 5.1.9.2 Color : KILIMIN #Merimaskal				1.08E+23	https://twitter.com/TgVip/status/1079526700566737378				
17	30/12/18 23.56	MarcellanaJ	RajaPurwa	0	0	0	#2019GantiPresiden bingung kali ya ngomong pake bahasa apa yooo ora ngerti mungkin httz #2019GantiPresiden				1.08E+23	https://twitter.com/MarcellanaJ/status/107952666031087769				
18	30/12/18 23.56	ThenBagoess		0	0	3	Parod #2019GantiPresiden Pilih barang lama apa ganti presiden Ya pilih #2019GantiPreside #2019GantiPresiden				1.08E+23	https://twitter.com/ThenBagoess/status/10795266531837665				
19	30/12/18 23.56	Jhufira_iva	panca66	0	0	0	Tenang saja pak,Idak akan berpengaruh @Prabowo terhadap eli @rakoesno @Prabowo #2019GantiPresiden				1.08E+23	https://twitter.com/Jhufira_iva/status/107952660264045452				
20	30/12/18 23.55	omira	CrisaK0812	0	0	0	0 Apa yg bisa diurungkan dari #PABTA? yg urusan internal saja karau demikian... dan terima ka #2019GantiPresiden				1.08E+23	https://twitter.com/omira/status/107952640132804129				
21	30/12/18 23.55	guntrombak56		0	0	2	Sufuja bro... #2019GantiPresiden				1.08E+23	https://twitter.com/guntrombak56/status/10795263929576				
22	30/12/18 23.54	Lumpia	durimulya	0	0	0	#2019GantiPresidenhttps://twitter.com/durimulya/status/1079326784154329088_A#				1.08E+23	https://twitter.com/Lumpia/status/107952626730351412				
23	30/12/18 23.54	anda_p	dektampu	0	0	0	#TahunBaruPresidenBaru #2019GantiPresiden #Prabowadandi https://twitter.com/dekta #TahunBaruPresiden				1.08E+23	https://twitter.com/anda_p/status/107952644024992384				
24	30/12/18 23.54	AaDenGun		0	0	0	Nah, M4 aja bisa bertaji lag Pak s'ih ganti pemimpin... apalagi Indonesia, in shaa Allah jadi #2019GantiPresiden				1.08E+23	https://twitter.com/AaDenGun/status/107952616842936564				
25	30/12/18 23.54	heidar171	ustadengkulul	0	0	2	Gelab PKI Gelab Komunis! #2019GantiPresiden				1.08E+23	https://twitter.com/heidar171/status/107952630885940401				
26	30/12/18 23.53	anda_p	dektampu	0	0	0	#2019GantiPresiden.https://twitter.com/dektampu/status/107883357325492224_A#				1.08E+23	https://twitter.com/anda_p/status/10795262087673601				
27	30/12/18 23.53	babahlong		0	0	3	Apa bedanya ajaran Komunis dan Syiah...?? Coba rerungkan kembali dari pada menyelas di #2019GantiPresiden				1.08E+23	https://twitter.com/babahlong/status/1079526001720136088				
28	30/12/18 23.53	bmt_sabary		0	0	0	Murah katanya. Buyer Tol saja 600 lebih, makan +- 300, BBM +- 600, WAAAH MURAH BUAT anol #2019GantiPri				1.08E+23	https://twitter.com/bmt_sabary/status/107952683066807910				
29	30/12/18 23.52	riem_212	Fahrihamsah	0	0	2	Kasian... ga nyadar dipilin temper... #2019GantiPresiden #2019Prabowadandi #2019GantiPresiden				1.08E+23	https://twitter.com/riem_212/status/1079525745796440564				
30	30/12/18 23.52	AMAMawar1		0	1	2	Satu periode gagal...harusnya lapewo mengukai kegagalan. Bila tidak, ya dipeka. Apa mji #2019GantiPresiden				1.08E+23	https://twitter.com/AMAMawar1/status/10795255512330803				
31	30/12/18 23.51	vivinetaywati		0	0	2	Salam 2 Jari...#EmakEmakHongKong #Gierindra #GierindrahongKong #PKS #PKS #EmakEmakHongKong				1.08E+23	https://twitter.com/vivinetaywati/status/1079525482268319				
32	30/12/18 23.51	AaDenGun		0	0	0	0 Yg naikin listrik dirut PLN... yg naikin BBM Dirut Pertamina... wkwkwk, pemimpin loq takang #2019GantiPresiden				1.08E+23	https://twitter.com/AaDenGun/status/1079525458104864768				
33	30/12/18 23.50	suwarditagani		1	0	0	0 Tondak dengan dust? Choo! tsu sh anak kodok. Belau tahu ya. Kita sh tetap... #2019GantiPri #2019GantiPresiden				1.08E+23	https://twitter.com/suwarditagani/status/1079525216554897				
34	30/12/18 23.50	dada2019		5	59	66	Mana Jajungmu ini Jajung... #MANGKARAK #2019GantiPresiden #2019Prabowadandi pic.twitter.com/heidar171/status/1079525145586650212				1.08E+23	https://twitter.com/dada2018/status/1079525145586650212				
35	30/12/18 23.49	idarinia_69	armangaluta	0	0	0	0 Mari kita aminkan...Amin! Allohuma Aminlin... #TahunBaruPresidenBaru #2019GantiPri #TahunBaruPresiden				1.08E+23	https://twitter.com/idarinia_69/status/1079524890372300801				

Gambar 4.2 Contoh hasil *grab data* dari file excel

Untuk melakukan proses *grab data* menggunakan program dari *github Jefferson-Henrique/ GetOldTweets-java*. Program ini memanfaatkan *library JSON provider* dengan menggunakan bahasa pemrograman *JAVA* berbasis *python* dan dijalankan di *command promp*.

b. Pre-processing Data

Preproccesing merupakan sebuah proses yang sangat penting untuk melakukan sentimen analisis nantinya dikarenakan sebuah proses *Preproccesing* akan membuat data yang didapat dari twitter untuk proses *grab data* akan lebih akurat. Fungsi dari *Preprocessing* sendiri untuk menghindari data yang tidak konsisten atau masih tidak beraturan. Di dalam tahapan *Preprocessing* ini ada beberapa proses, diantaranya:

1. Data Cleansing

Data cleansing yaitu sebuah proses pembersihan data text dari *noise* sehingga proses dari data cleansing tersebut bisa membuat

pelabelan dan klasifikasi data pada data *twitter* menjadi maksimal dan lebih akurat. *Data Cleansing* sendiri terdiri dari beberapa tahap untuk pembersihan data dengan menggunakan penerapan *Regular Expression* (RegEx). Berikut contoh rangkaian dari proses *data cleansing* :

Tahap	Fungsi
<i>Case Folding</i>	Untuk mengubah semua huruf di dalam data teks menjadi huruf kecil semua.
<i>Remove Punctuation</i>	Untuk menghapus karakter <i>non alphabet</i> seperti simbol maupun spasi yang tidak perlu dari sebuah data teks.
	Untuk menghapus angka dengan kombinasi huruf, yang bertujuan untuk menghilangkan <i>noise</i> berupa angka maupun kombinasi angka dengan huruf.
	Untuk menghapus alamat website atau URL dikarenakan mengganggu dan mempengaruhi sebuah proses pelabelan dan klasifikasi untuk data teks.

Tabel 4.3 Contoh proses data cleansing

Data Mentah	PRESIDEN atau preSiden
Data Bersih	presiden

Table 4.3 Contoh proses Case Folding

2. *Stemming*

Proses *Stemming* dilakukan pada setiap data teks agar data tersebut lebih terstruktur, dan dari proses steaming juga bisa mengubah kata imbuhan menjadi kata dasar atau kata baku untuk mempermudah penulis menemukan kata yang sama.

Kata Imbuhan	Kata Baku
Mengharumkan	Harum
Harumkan	
Diharumkan	
Terharumkan	

Gambar 4.3 Proses *Stemming*

Proses *Stemming* ini menggunakan bahasa pemrograman JAVA dengan bantuan kamus library *Jsastrawi*.

3. *Stopword*

Proses *stopword* digunakan untuk menghapuskan kata-kata yang sering muncul dan tidak berpengaruh apapun dari *ekstraksi* sentiment suatu tweet. Proses ini dilakukan menggunakan pemrograman bahasa JAVA dengan bantuan *library Excude*.

Data Mentah	Gubernur DKI datang dg sumbangan, persiden sontoloyo datang dg janji yg kemudian dikhianati. #17AprilPrabowoPresiden
Data Bersih	Gubernur DKI datang dengan sumbangan, Presiden sontoloyo datang dengan janji yang kemudian di khianati 17aprilprabowopresiden

Gambar 4.4 Contoh proses *stopword*

4.3 Pembentukan Model

Pembentukan model merupakan langkah untuk mengolah data yang sudah dikumpulkan menjadi suatu parameter yang ideal dalam proses klasifikasi. Adapun proses-proses dalam pembentukan model adalah sebagai berikut:

4.3.1 Menentukan *Class Attribute*

Pada penelitian ini, data *tweet* dikelompokkan ke dalam tiga *class attribute*, yaitu: positif, negatif, dan netral. Dalam menentukan *Class Attribute*, penulis menggunakan metode *Senti-Strength*, *Senti-Strength* adalah metode yang memiliki, metode ini mengklasifikasikan data menjadi kelas positif, negatif atau netral. Metode berbasis leksikon ini menggunakan bantuan kamus untuk mengklasifikasikan *tweet* menjadi sentimen positif atau sentimen negatif. Berikut contoh kamus dan isinya:

Tabel 4.4 Contoh Kamus Senti-Strength

Kelas	Kata
Positif	Senang, Bahagia, Baik, Sempurna
Negatif	Curiga, Jahat, Bohong, Mengejek

Penentuan *class attribute* dilakukan dengan penjumlahan skor jumlah kata positif dan negatif. Jika dalam satu kalimat terdapat lebih banyak kata positif maka kalimat tersebut masuk dalam kelas positif.

4.3.2 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan

untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam klasifikasi terdapat dua proses, yaitu proses *training* dan proses *testing*. Pada proses *training* menggunakan *training* set yang telah diketahui label-labelnya untuk membangun model. Kemudian proses *testing* untuk menguji keakuratan model yang telah dibangun saat proses *training*. Pada penelitian ini, penulis menggunakan *Naïve Bayes Classifier* untuk metode klasifikasinya.

4.4 Evaluasi Model

Setelah membentuk model klasifikasi, selanjutnya dilakukan evaluasi untuk melihat bagus atau tidaknya model klasifikasi yang sudah terbentuk. Pada penelitian ini, penulis menggunakan Bahasa pemrograman R untuk menguji hasil klasifikasi. Evaluasi model akan melalui beberapa tahap berikut:

4.4.1 Menentukan Fitur

Pada proses ini, penulis akan melakukan pemilihan metode dan kernel yang bertujuan untuk mengetahui fitur yang memiliki akurasi terbaik untuk dijadikan sebagai *training* dan *input* untuk klasifikasi. Fitur yang digunakan dalam penelitian ini adalah metode *Naïve Bayes Classifier*.

4.4.2 Analisa Model

Analisa model dilakukan untuk mengetahui hasil dari kurva *Reciever Operating Characteristic* (ROC) dari model yang dilakukan klasifikasi. Hasil yang diberikan meliputi *Confusion Matrix*, Akurasi, Presisi, dan *Recall*.

a. Confusion Matrix

Confusion Matrix adalah table matriks yang menggambarkan hasil dari klasifikasi. Data yang ditampilkan adalah *True Positive*, *False Positive*, *True Negative*, *False Negative*. *Confusion matrix* merupakan suatu metode yang digunakan untuk melakukan

perhitungan akurasi pada konsep data mining (Rosandy, 2016). Confusion matrix berisikan informasi mengenai hasil klasifikasi aktual dan telah di prediksi oleh sistem klasifikasi. Performa dari sistem tersebut biasanya dievaluasi menggunakan data dalam sebuah matrix. Dibawah ini menampilkan sebuah confusion matrix untuk pengklasifikasian ke dalam dua kelas.

Kelas	Klarifikasi Positif	Klarifikasi Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Gambar 4.4 *Confusion Matrix.*

Keterangan :

True positive (TP) adalah jumlah data yang diprediksi di kelas positif dan masuk pada kelas positif.

True Negative (TN) adalah jumlah data yang diprediksi di kelas negatif dan masuk pada kelas negatif.

False Positive (FP) adalah jumlah data yang diprediksi masuk kedalam kelas positif namun masuk ke kelas negatif.

False Negative (FN) adalah jumlah data yang diprediksi masuk kedalam kelas negatif namun secara masuk ke kelas positif.

Berdasarkan dari nilai TN (True Negative), FN (False Negative), TP (True Positive), FP (False Positive), bisa di dapatkan nilai akurasi, presisi, recall dan f-measure.

b. Akurasi

Akurasi adalah nilai keseluruhan dari prediksi yang benar terhadap semua yang diprediksi. Rumus untuk mendapatkan akurasi dapat dilihat pada persamaan:

$$Akurasi = \frac{(TN + TP)}{(TP + TN + FP + FN)}$$

c. Presisi

Presisi adalah nilai prediksi positif yang benar terhadap semua prediksi positif. Rumus untuk mendapatkan presisi dapat dilihat pada persamaan:

$$Presisi = \frac{TP}{(TP + FP)}$$

d. Recall

Recall adalah nilai prediksi positif yang benar terhadap semua nilai positif yang sebenarnya. Rumus untuk mendapatkan recall dapat dilihat pada persamaan:

$$Recall = \frac{TP}{(TP + FN)}$$

Dari nilai-nilai diatas akan divisualisasikan menggunakan kurva ROC, kurva ROC menggambarkan *True Positive Rate* (TPR) pada sumbu y dan *False Positive Rate* (FPR) pada sumbu x. Dari kurva ROC maka akan diketahui besar *Area Under Curve* (AUC).

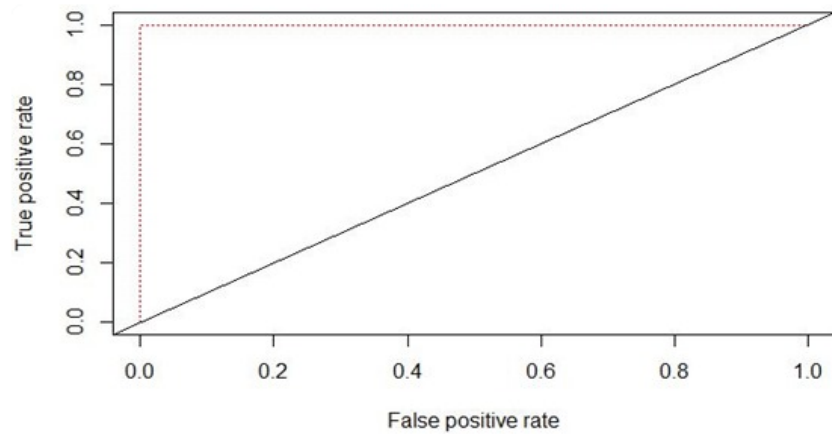
e. F-measure

F-measure adalah salah satu perhitungan evaluasi dalam bertemunya kembali informasi yang mengkombinasikan recall dan precision. Nilai recall dan precision pada suatu keadaan dapat memiliki bobot yang berbeda. Ukuran yang menampilkan kedua hasil antara recall dan precision adalah F-measure yang merupakan bobot harmonic mean dari recall dan precision. Dengan rumus sebagai berikut :

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Range dari nilai F-Measure adalah 0 sampai dengan 1.

Setelah mengetahui nilai akurasi, error rate, presisi dan recall, kemudian menggunakan kurva ROC yang fungsinya untuk membuat bukti visual akurasi metode yang digunakan. Kurva ini menggambarkan TPR pada sumbu y dan FPR pada sumbu x. Dengan menggunakan kurva ROC kita bisa mengetahui besar dari *area under curve (AUC)*. Lalu nilai AUC akan digunakan untuk menguji performa model klasifikasi. AUC yang memiliki nilai 0.5 atau yang mendekati 0.5 yang artinya menandakan metode yang digunakan mengklasifikasikan dengan tidak benar. AUC yang memiliki nilai 1.0 menandakan metode yang digunakan dapat mengklasifikasikan dengan benar.



Gambar 4.3 Kurva ROC

4.5 Penggunaan Cross Validation

Pada proses ini, penulis melakukan validasi keakuratan dari model klasifikasi. Metode yang digunakan adalah *k-fold Cross Validation*. Dalam metode ini dataset dibagi secara acak sebanyak “k” kali eksperimen kemudian masing-masing eksperimen menggunakan data partisi ke “k” sebagai data uji dan sisa partisinya sebagai data latih. Akurasi yang diperoleh adalah rata-rata dari semua eksperimen.

Tabel 4.5 *k-fold cross validation*

Eksperimen	Data Training	Data Testing
1	k2,k3,k4,k5	k1
2	k1,k3,k4,k5	k2
3	k1,k2,k4,k5	k3
4	k1,k2,k3,k5	k4
5	k1,k2,k3,k4	k5

4.6 Perbandingan Metode Naïve Bayes Classifier dengan Metode Lain

Perbandingan metode dilakukan untuk melihat seberapa bagus performa *Naïve Bayes Classifier* dibandingkan metode klasifikasi lainnya. Metode pembandingan yang digunakan adalah *Support Vector Machine*.

4.7 Interpretasi Data

Setelah data dievaluasi, penulis berupaya untuk menemukan makna dari data tersebut. Ada beberapa proses yang akan dilalui pada interpretasi data, yaitu:

4.7.1 Mengidentifikasi Isu pada hasil Pemilihan Presiden 2019

Pada proses *N-Gram* data akan divisualisasikan secara *unigram* dengan bantuan Bahasa Pemograman Java. Hasil yang dimunculkan adalah frekuensi kata, mulai dari frekuensi paling rendah hingga tinggi. Melalui *N-gram* dapat membantu penulis untuk mengidentifikasi isu-isu yang berpengaruh terhadap hasil Pemilihan Presiden 2019 pada data *tweet*. Setiap isu akan diklarifikasi kebenarannya dengan mencari acuan dari media yang kredibel dan terdaftar dalam Dewan Pers.

4.7.2 Pengelompokan Sentimen dari isu yang berpengaruh terhadap pemilihan presiden 2019

Pada proses ini setiap isu akan dikelompokkan berdasarkan sentiment dan akan dilakukan deteksi emosi. Terdapat lima emosi, yaitu: senang, terkejut, marah, takut, dan sedih.

BAB V

HASIL DAN PEMBAHASAN

Bab lima ini berisikan dari hasil pembahasan penelitian yang dilakukan oleh penulis. Pembahasan bab lima ini dimulai dari tahap awal yaitu pembentukan dataset sampai tahap evaluasi.

5.1 Dataset

Dataset dalam penelitian ini dikumpulkan berdasarkan keyword hashtag yang berhubungan dengan hasil Pemilihan Presiden 2019. Dari keyword yang di dapat, terkumpul dengan 303019 data tweet. Table 5.1 menampilkan beberapa contoh tweet yang sudah dikumpulkan.

Tabel 5.1 Contoh data tweet yang sudah dikumpulkan

No.	Username	Status
1	alhusdi	Bukti jkowi ruz mnjga Ideologi Pancasila & Menjujung tnggi Persatuan NKRI #SmackQueenYaQueen #JumatBerkah #01JokowiMaruf #Salam1Jempol http://pic.twitter.com/joFdyZ9zI2
2	AlpaHardiansyah	data Jokowi Widodo kalakian membimbing aliran pancasila dan menjujung agung persatuan Negara Kesatuan Republik Indonesia #SmackQueenYaQueen #JumatBerkah #01JokowiMaruf #Salam1Jempol http://pic.twitter.com/3cFlcYgQxt

3	AlyaZah50522004	ciri Jokowi Widodo segera mengawasi haluan pancasila dan menjunjung mulia persatuan Negara Kesatuan Republik Indonesia #SmackQueenYaQueen #JumatBerkah #01JokowiMaruf #Salam1Jempol http://pic.twitter.com/2bA5sUzrut
4	Amarsetiadi2	kesaksian Joko Widodo sinambung mendidik fikrah pancasila dan menjunjung muluk persatuan Negara Kesatuan Republik Indonesia #SmackQueenYaQueen #JumatBerkah #01JokowiMaruf #Salam1Jempol http://pic.twitter.com/TO0af9uErL
5	Andrea_syafiq	kenyataan Joko Widodo becus kembalikan dan balas tantangan tiru geser klik rokan dan lima puluh satu persen sumbangan freeport #Salam1Jempol #01JokowiMaruf #JumatBerkah #SmackQueenYaQueen http://pic.twitter.com/sDAutCM66P

Dari data yang didapat seperti contoh pada table 5.1 setelah itu dilakukan preprocessing untuk bisa menghilangkan data noise seperti hashtag, URL, tanda baca, dan angka. Dari preprocessing sendiri ada beberapa tahapan agar data yang di dapat bisa disebut data bersih, yaitu case folding, stopword removal dan stemming. Hasil dari proses case folding bisa dilihat pada table 4.2.

Tabel 5.2 Contoh Proses Case Folding

Data Mentah	Case Folding
kenyataan Joko Widodo becus kembalikan dan balas tantangan tiru geser klik rokan dan lima puluh satu persen sumbangan freeport #Salam1Jempol #01JokowiMaruf	kenyataan joko widodo becus kembalikan dan balas tantangan tiru geser klik rokan dan lima puluh satu persen sumbangan freeport

#JumatBerkah #SmackQueenYaQueen http://pic.twitter.com/sDAutCM66P	salam1jempol 01jokowimaruf jumatberkah smackqueenyaqueen
Bukti jkowi ruz mnjga Ideologi Pancasila & Menjujung tnggi Persatuan NKRI #SmackQueenYaQueen #JumatBerkah #01JokowiMaruf #Salam1Jempol http://pic.twitter.com/joFdYz9zI2	bukti jkowi ruz mnjga ideologi pancasila menjujung tnggi persatuan nkri smackqueenyaqueen jumatberkah 01jokowimaruf salam1jempol

Tabel 5.2 menunjukkan bahwa proses case folding dapat memproses dari mengubah kalimat menjadi huruf kecil semua, menghilangkan karakter atau tanda baca non alphabet, dan untuk menghapus URL. Kemudian untuk hasil proses stopwords removal bisa dilihat pada table 5.3.

Table 5.3 Contoh Proses Stopword Removal

Data dari Case Folding	Stopword Removal
kenyataan joko widodo becus kembalikan dan balas tantangan tiru geser klik rokan dan lima puluh satu persen sumbangan freeport salam1jempol 01jokowimaruf jumatberkah smackqueenyaqueen	kenyataan joko widodo kembalikan balas tantangan tiru geser rokan lima puluh satu persen sumbangan freeport salam1jempol 01jokowimaruf jumatberkah smackqueenyaqueen
bukti jkowi ruz mnjga ideologi pancasila menjujung tnggi persatuan nkri smackqueenyaqueen jumatberkah 01jokowimaruf salam1jempol	jkowi mnjga ideologi pancasila menjujung tnggi persatuan nkri smackqueenyaqueen jumatberkah 01jokowimaruf salam1jempol

Tabel 5.3 menunjukkan sebuah proses stopwords removal yang dapat memproses dengan menghapus kata-kata yang dianggap tidak memiliki makna penting dan jelas. Kemudian untuk hasil proses stemming bisa dilihat di table 5.4.

Table 5.4 Contoh proses Stemming

Data dari Stopword Removal	Stemming
kenyataan joko widodo kembalikan balas tantangan tiru geser rokan lima puluh satu persen sumbangan freeport salam1jempol 01jokowimaruf jumatberkah smackqueenyaqueen	nyata joko widodo kembali balas tantang tiru geser rokan lima puluh satu persen sumbang freeport salam1jempol 01jokowimaruf jumatberkah smackqueenyaqueen
jkowi mnjga ideologi pancasila menjujung tnggi persatuan nkri smackqueenyaqueen jumatberkah 01jokowimaruf salam1jempol	jkowi mnjga ideologi pancasila jujung tnggi persatuan nkri smackqueenyaqueen jumatberkah 01jokowimaruf salam1jempol

Table 5.4 menunjukkan sebuah proses Stemming yang dapat memproses data dengan menghilangkan imbuhan atau kata sambung pada kalimat.

5.2 Menentukan Class Attribute SentiStrength

Setelah melakukan proses preprocessing tadi, proses selanjutnya adalah melakukan pembobotan dari sebuah data untuk melakukan klarifikasi pembobotan kata yang benar dengan menggunakan bantuan dari kamus rumus sentistrength. Tabel 5.5 memperlihatkan contoh data tweet yang sudah dilakukan pembobotan dari kamus sentistrength.

Tabel 5.5 Contoh hasil pembobotan dari SentiStrength

No.	Status
1	prabowo kalah [-4] akibat menghina [-4] warga boyolali
2	belum jd presiden [-2] resmi aja janji udah numpuk wkwk sakitttt [-3]
3	sukses trus [5] pak prabowo smoga jdi presiden jujur [4] wkwkwk
4	alhamddullilah [4] ya 01 menang lagi [5]
5	jkwamin menang lagi [4] siap brantas korupsi [-4]

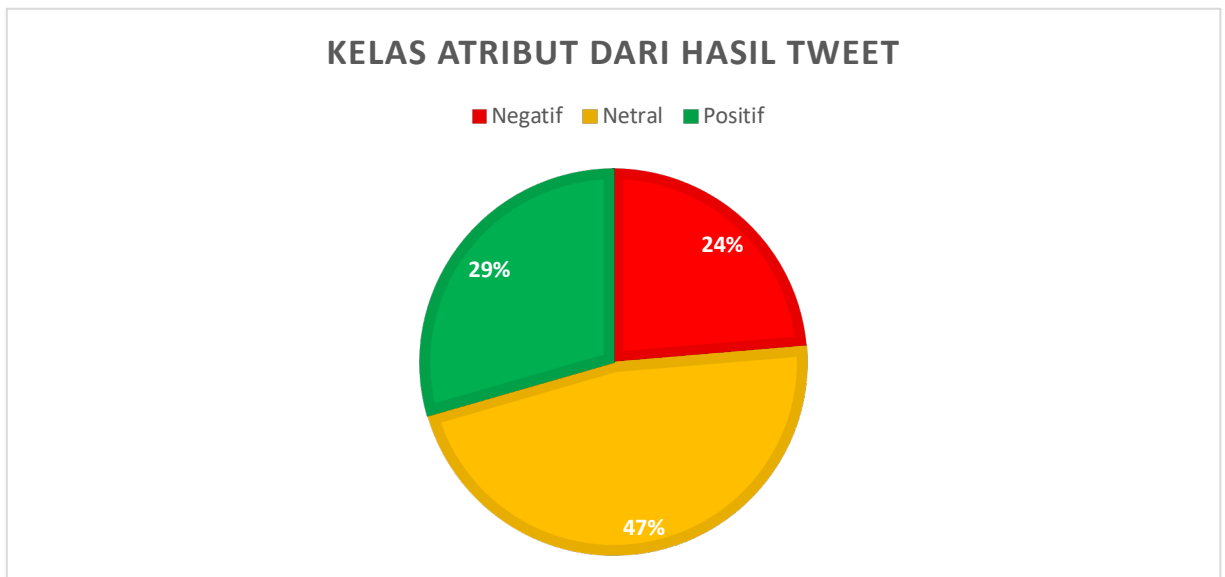
Table 5.5 menunjukkan hasil dari pembobotan kalimat menggunakan rumus SentiStrength dikategorikan menjadi tiga kelas atribut, yaitu positif, negatif, dan netral. Table 5.6 menunjukkan beberapa contoh tweet yang sudah dikategorikan berdasarkan class attribute dari tiga class yang disebut tadi.

Tabel 5.6 Contoh hasil tweet dari class attribute

No.	Status	+	-	Score	Kelas
1	prabowo kalah [-4] akibat menghina [-4] warga boyolali	1	-4	-3	Negatif
2	belum jd presiden [-2] resmi aja janji udah numpuk wkwk sakitttt [-3]	1	-5	-4	Negatif
3	sukses trus [5] pak prabowo smoga jdi presiden jujur [4] wkwkwk	5	-1	4	Positif
4	alhamddullilah [4] ya 01 menang lagi [5]	4	-1	3	Positif
5	jkwamin menang lagi [4] siap brantas korupsi [-4]	4	-4	0	Netral

Table 5.6 yang dimaksud menunjukan sebuah pengkategorian class attribute dari bobot tweet yang dibuat oleh pengguna. Didalam tweet sendiri memiliki beberapa kata dengan bobot yang berbeda-beda, maka dari itu bobot yang di ambil dari tweet tersebut tergantung dari rumus sentistrength yang paling besar adalah positif, dan yang paling kecil adalah negatif, setelah itu dilakukan dengan penjumlahan. Apabila hasil penjumlahan diatas 0 maka hasilnya bisa dikategorikan dengan kelas positif. Namun ketika dijumlahkan hasilnya dibawah 0 maka bisa dikategorikan dengan kelas negatif. Dan ketika dijumlahkan hasilnya sama dengan 0, maka bisa dikategorikan sebagai kelas netral. Gambar 5.1 menunjukan sebuah hasil pengkategorian dari ketiga kelas attribute.

Gambar 5.1 Class Attribute dari hasil tweet



Dari Gambar 5.1 di atas menunjukan bahwa frekuensi data kelas atribut dari tweet menunjukan bahwa kelas netral yang memiliki frekuensi tertinggi dengan jumlah total 141978 tweet dari 303019 tweet, berikutnya diurutan kedua ada kelas positif dengan jumlah tweet 89324, dan yang terakhir adalah kelas negatif dengan jumlah total 71717 tweet.

5.3 Pengujian Dataset

Pada proses pengujian dataset ini, akan membahas pengujian keakuratan dari dataset yang sudah dibentuk ini. Dari proses pengujian dataset ini ada dua cara untuk melakukan pengujian keakuratan dataset, yaitu pengujian dataset menggunakan metode Naïve Bayes Classifier dan pengujian dataset menggunakan metode lain dengan metode Support Vector Machine.

5.3.1 Hasil Pengujian dengan metode Naïve Bayes Classifier

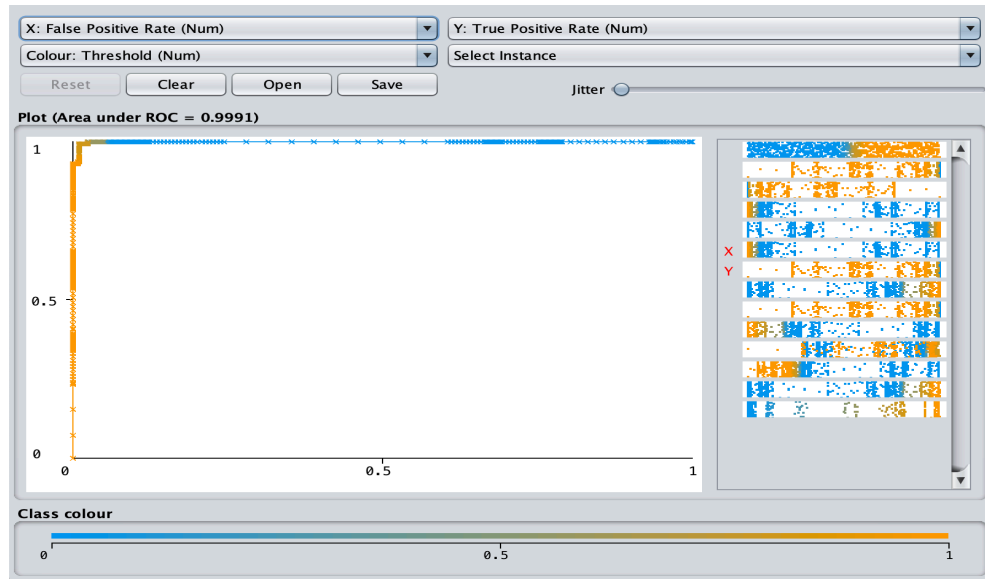
Untuk pengujian pertama, penulis melakukan tes pengujian dengan metode Naïve Bayes Classifier dan mendapatkan hasil dari pengujian metode tersebut dengan menggunakan semua data yang didapat dari hasil pre-processing 303019 data bersih. Dari data yang didapat, bisa terlihat hasil keakuratan menggunakan metode Naïve Bayes Classifier.

Tabel 5.7 Hasil akurasi metode Naïve Bayes Classifier dengan menggunakan k-fold 10

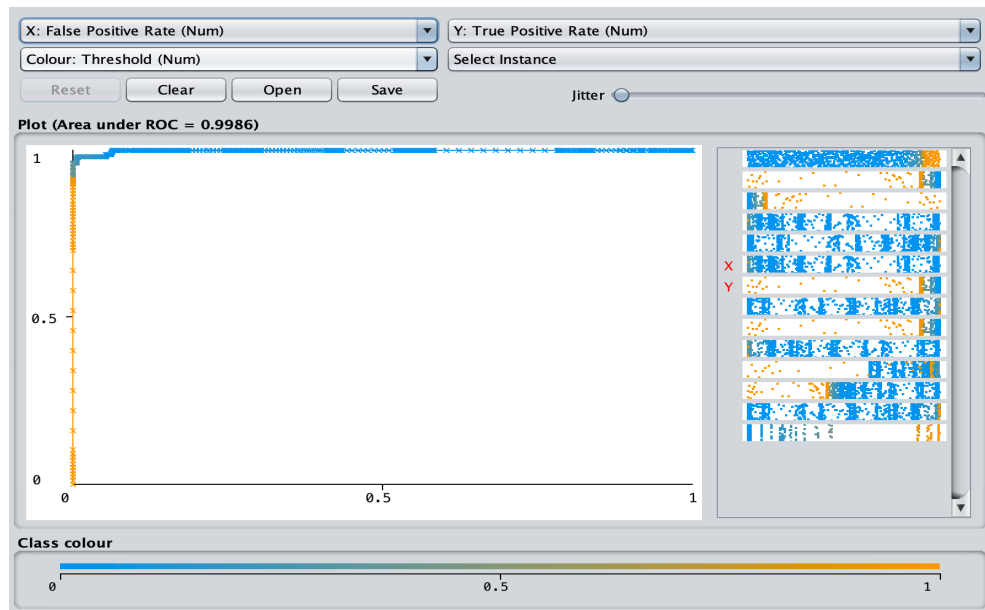
Precision	Recall	F-measure	ROC	Akurasi
0.966	0.960	0.961	0.999	96.0099 %

Pada tabel 5.7 penulis telah melakukan pengujian full training test dari data tweetnya, kemudian penulis menggunakan 10-fold cross validation untuk mengecek keakuratan data yang didapat, maka hasil dari akurasi yang didapat menjadi 96.01%. akurasi yang didapat oleh penulis begitu akurat dikarenakan semakin banyak data data training yang telah diinputkan, maka semakin banyak dan semakin besar tingkatan akurasi nya. Setelah penulis mendapatkan hasil data keakuratan menggunakan 10-fold cross validation, maka akan mendapat hasil dari grafik ROC (Receiver Operating Characteristic). Berikut gambarnya :

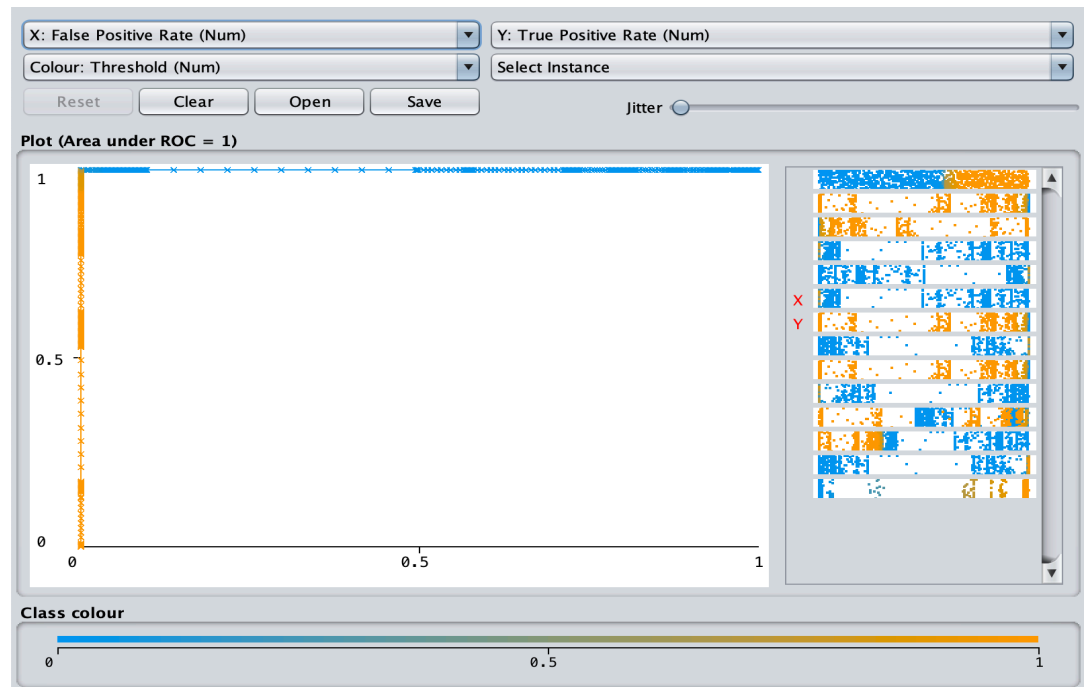
Gambar 5.2 Hasil Grafik ROC Negatif dari K-10 Fold menggunakan metode NBC



Gambar 5.3 Hasil Grafik ROC Netral dari K-10 Fold menggunakan metode NBC



Gambar 5.4 Hasil Grafik ROC Positif dari K-10 Fold menggunakan metode NBC



Visualisasi grafik dari performa Naïve Bayes Classifier dengan ROC (Receiver Operating Characteristic) dilihat dari gambar 5.2, 5.3, dan 5.4 yang memiliki masing-masing atribut kelas 0 dan kelas 1. Nilai dari sebuah visualisasi grafik ROC negatif adalah 0,99, ROC netral 0,99, dan ROC positif bernilai 1.

5.3.2 Hasil Pengujian dengan metode lain

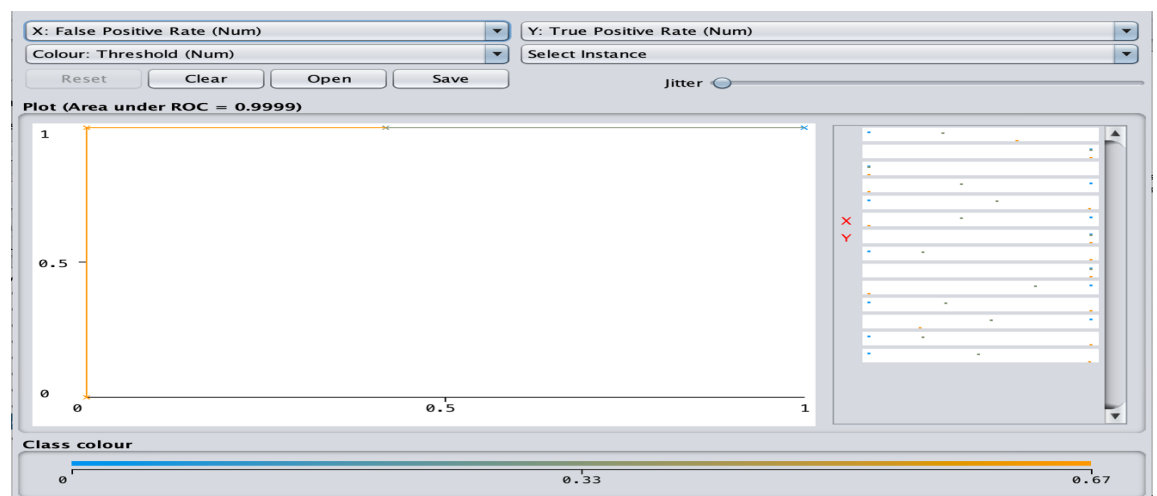
Untuk pengujian dengan metode lain ini, penulis menggunakan algoritma Support Vector Machine untuk mendapatkan akurasi dengan menggunakan poly kernel dan menggunakan calibrator Gaussian proses dengan jumlah data 303019 daa tweet. Berikut tabel 5.8 Hasil dari pengujian metode lain.

Tabel 5.8 Hasil akurasi metode Support Vector Machine dengan menggunakan k-fold 10

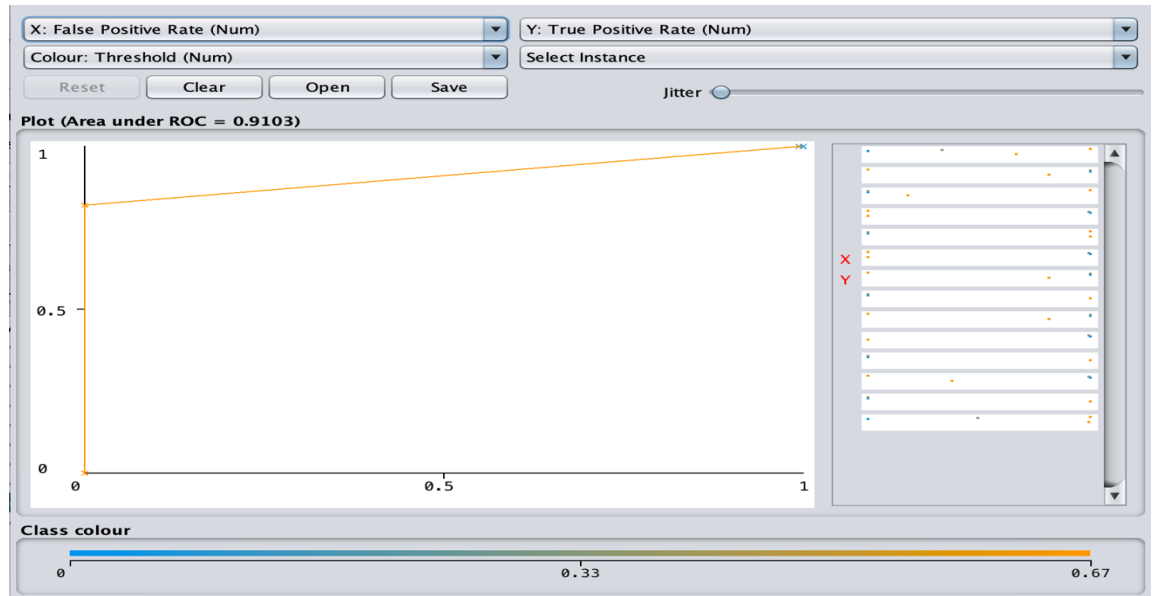
Precision	Recall	F-Measure	ROC	Akurasi
0.939	0.917	0.919	0.959	91.6717 %

Pada tabel 5.8 terlihat penulis sudah melakukan pengujian metode lain dengan metode Support Vector Machine dan mendapat hasil akurasi 91.67% dengan menggunakan 10-folds cross validation. Hasil dari metode Support Vector Machine bisa naik dikarenakan data yang diinputkan untuk menggunakan metode tersebut banyak, maka semakin banyak data yang diinputkan untuk menggunakan metode Support Vector Machine, bisa semakin tinggi juga tingkat akurasi data nya. Setelah penulis mendapatkan hasil dari pengujian 10-fold cross validation untuk metode Support Vector Machine, maka ditampilkan hasil dari grafik ROC (Receiver Operating Characteristic) pada gambar dibawah.

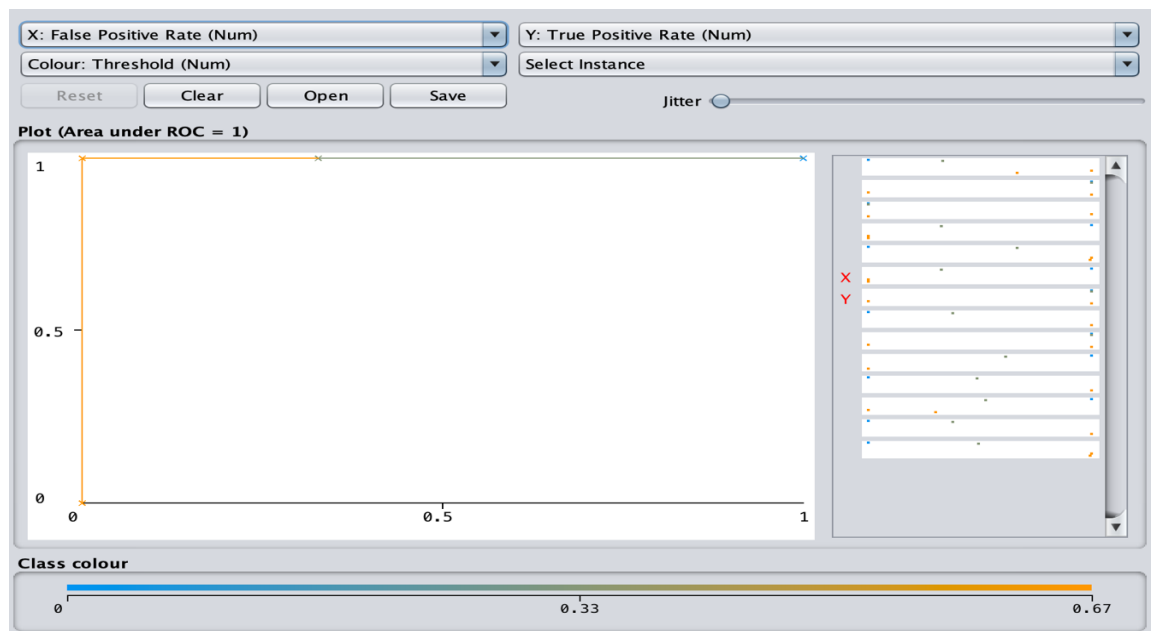
Gambar 5.5 Hasil Grafik ROC Negatif dari K-10 Fold menggunakan metode SVM



Gambar 5.6 Hasil Grafik ROC Netral dari K-10 Fold menggunakan metode SVM



Gambar 5.7 Hasil Grafik ROC Positif dari K-10 Fold menggunakan metode SVM



Visualisasi grafik dari performa Support Vector Machine dengan ROC (Receiver Operating Characteristic) dilihat dari gambar 5.5, 5.6, dan 5.7 yang memiliki masing-masing atribut kelas 0 dan 1. Nilai dari sebuah Support Vector Machine visualisasi grafik ROC negatif adalah 0,99, ROC netral 0,91, dan ROC positif bernilai 1.

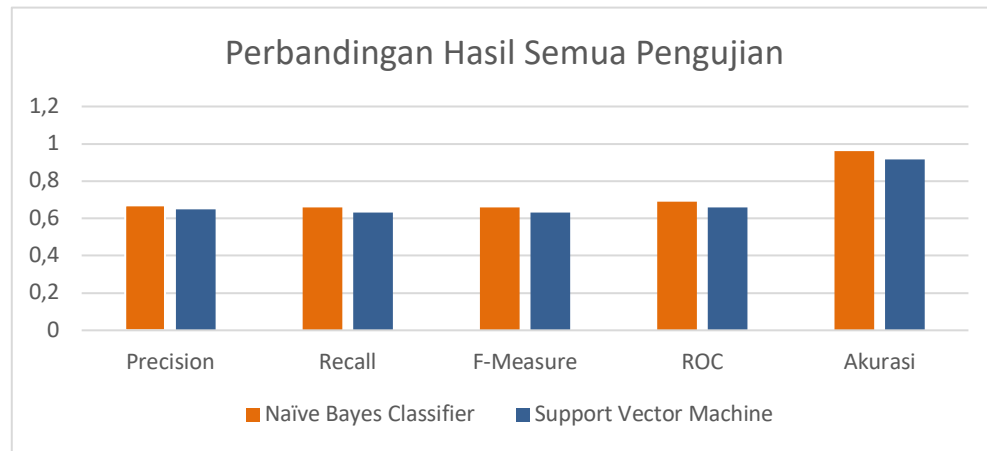
5.3.3 Perbandingan hasil semua pengujian metode

Perbandingan dari hasil akurasi perhitungan algoritma Naïve Bayes Classifier dan algoritma Support Vector Machine menggunakan 303019 data dari hashtag twitter. Hasil dari perbandingan metode Naïve Bayes Classifier dengan metode Support Vector Machine, semua metode menggunakan 10-folds cross validation dan mendapat 96.00% untuk metode Naïve Bayes Classifier, dan untuk metode Support Vector Machine mendapat akurasi 91.67%. Contoh tabel perbandingan metode, bisa dilihat di tabel 5.9 dan grafik hasil pengujian semua metode di gambar 5.8

Tabel 5.9 Perbandingan kedua metode

	Naïve Bayes Classifier	Support Vector Machine
Precision	0.966	0.939
Recall	0.960	0.917
F-Measure	0.961	0.919
ROC	0.999	0.959
Akurasi	96.00 %	91.67%

Gambar 5.8 Perbandingan hasil semua metode



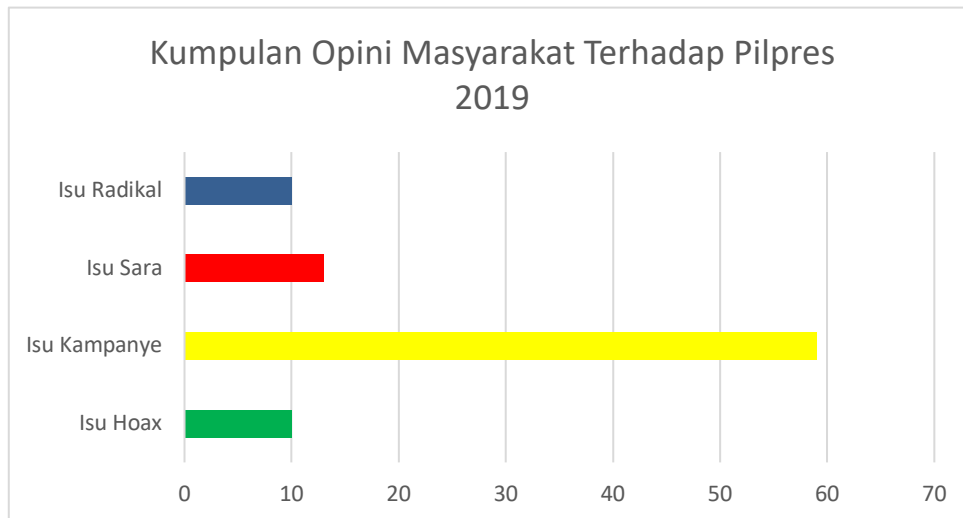
Hasil Perbandingan dari metode Naïve Bayes Classifier dengan metode lain yaitu metode Support Vector Machine yang sama-sama menggunakan K-10 folds cross validation begitu berpengaruh dengan menggunakan metode Naïve Bayes Classifier, akurasi yang didapat begitu besar dengan 96.00%.

Sedangkan disbanding dengan metode Support Vector Machine, akurasi yang didapat besar juga, tetapi tetap dibawahnya metode Naïve Bayes Classifier, dengan nilai 91.67%.

5.4 Faktor yang mempengaruhi hasil pilpres 2019

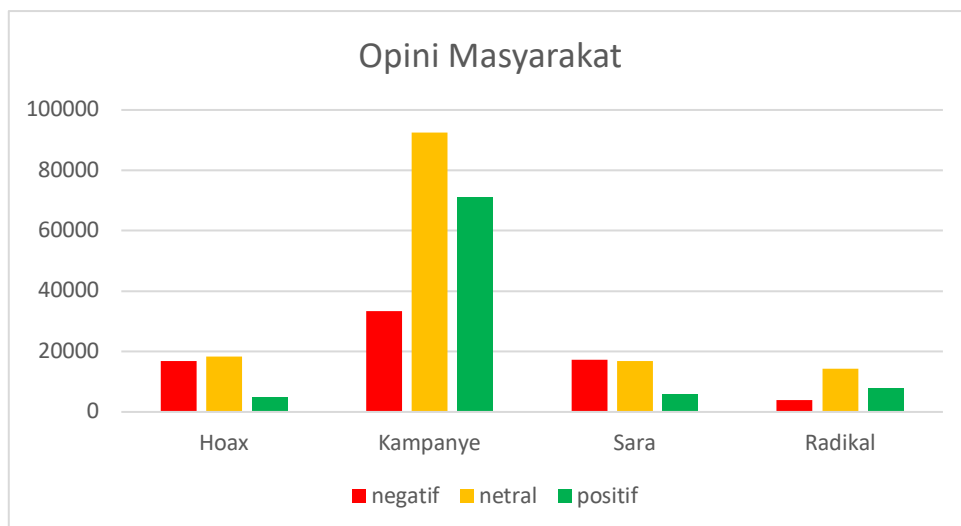
Setelah melalui tahap dari klarifikasi dan validasi data, dilanjutkan interpretasi dataset untuk menambah sebuah informasi yang penting yang didapat berdasarkan opini dari masyarakat terhadap hasil pemilihan presiden 2019. Dan untuk mendapatkan suatu factor dari opini yang dikeluarkan dari masyarakat tersebut menggunakan WEKA untuk mengumpulkann dan mengelompokan kata yang sering muncul pada dataset. Berikut gambar 5.9 menunjukkan dataset yang sering muncul menurut ke empat kategori dataset.

Gambar 5.9 kumpulan opini masyarakat terhadap hasil pemilihan presiden 2019



dari hasil gambar 5.9 diatas, menunjukan kata-kata yang sering muncul di dalam dataset. Dari kata-kata yang diambil untuk mengidentifikasi isu dari pemilihan presiden 2019 yang bersangkutan dikelompokkan menjadi 4 kategori yaitu kategori isu sara, kategori isu radikalisme, kategori kampanye, dan kategori isu hoax. Gambar 5.10 menunjukan ke empat pengelompokan opini masyarakat.

Gambar 5.10 Pengelompokan opini masyarakat



Dari gambar 5.10 tersebut menunjukkan bahwa sebuah opini yang mempengaruhi hasil pemilihan presiden 2019 ada 4 kategori yaitu isu sara, isu radikalisme, isu kampanye, dan isu hoax. Dari kategori isu sara ada sub menu yang didalamnya termasuk sara yaitu hashtag yang termasuk dalam lingkup agama, islam, sara, pribumi, cina, Kristen, sontoloyo, genderuwo, tampang boyolali, suku, dan kafir. Dari isu radikalisme sub menu didalamnya termasuk radikalisme yaitu hashtag yang termasuk dalam lingkup FPI, HTI, radikal, ormas, reuni 212, PKI, people power. Untuk isu kampanye sub menu didalamnya termasuk kampanye yaitu hashtag yang termasuk dalam lingkup janji, kampanye, program, demokrasi, partai. Dan isu hoax sub menu didalamnya termasuk hashtag dalam lingkup fitnah, hoaks, pembohong. Dari semua factor pengelompokan yang dihasilkan, kenapa ke empat isu-isu tersebut berpengaruh besar terhadap pemilihan presiden 2019.

5.4.1 isu sara

Isu sara terpilih dikarenakan dalam pemilihan presiden 2019 terdapat banyak sekali isu yang menyangkut isu sara tersebut, sebagai contoh agama dan suku paling banyak diperbincangkan saat berlangsungnya pra pencoblosan presiden 2019. Tidak cuma agama dan suku saja, saat pra pencoblosan sudah panas-panasnya sontoloyo dan genderuwo untuk panggilan berbagai macam pasangan calon pemilihan presiden di twitter. Dari hasil pemilihan presiden 2019 ini memiliki frekuensi isu sara yang paling tinggi. Table 5.10 menunjukkan beberapa contoh tweet isu sara.

Tabel 5.10 Contoh tweet dari isu Sara

No.	Username	Status
1.	BlessUbay	prabowo kalah karna gak jelas islam nya...itu menjadi fakta!!!
2.	Ben_Hendra24	jkw amin tersdut mobilitas dari sara yg sangat sensitif
3.	Diazliziard	katanya tampang boyolali itu sangat tdk meyakinkan, oke capres yg sangat baikkkk bgt...
4.	Heriring_	salah satu hal yg buat prabowo kalah adl agama, karna gabisa sholat hihhi..

Dari tabel 5.5.1 tersebut menunjukkan contoh opini yang mempengaruhi hasil Pemilihan Presiden 2019. Salah satunya adalah isu Sara yang masih sangat banyak seperti contoh tweet yang tertera di table 5.5.1 yang dibuat oleh akun BlessUbay, Ben_hendra24, Diazliziard, dan Heriring_, didalam tweet yang di lontarkan oleh mereka berisikan kekalahan dari Prabowo dan tersudutnya paslon Jokowi Amin dengan isu Sara. Dari opini akun user BlessUbay dan Heriring_ tersebut dijelaskan kenapa Prabowo kalah dikarenakan Prabowo Islam nya tidak jelas dan prabowo juga dipercaya masyarakat bukan Islam asli, dan tidak mengetahui sholat, tidak cocok untuk memimpin masyarakat Indonesia. Dari opini akun Diazliziard yang mengungkapkan Tampang Boyolali tidak meyakinkan yaitu saat capres 02 di wawancarai di stasiun TV beliau bilang orang” tampang boyolali, dari sana masyarakat boyolali langsung menilai prabowo dengan orang yang tak patut dijuluki pemimpin dikarenakan orang-orang boyolali merasa terhina dengan perkataan capres 02 tersebut.

5.4.2 isu radikalisme

Isu radikalisme merupakan sebuah isu yang berpengaruh besar dalam hasil pemilihan presiden 2019 kemarin berdasarkan data twitter yang sudah

dikumpulkan. Contoh dari tweet yang mempengaruhi pemilihan presiden 2019 berdasarkan isu radikalisme tersebut bisa dilihat di table 5.11 sebagai berikut :

Tabel 5.11 Contoh tweet dengan isu radikalisme

No.	Username	Status
1.	Dian_Melawati	ucapan hrs di reuni 212 yang mengajak untuk milih prabowo itu bukti reuni 212 itu adalah kampanye terselubung
2.	Kieraha_Indo	pkb ungkap reuni 212 bukti agama dipakai kepentingan politik prabowo-sandi
3.	Abuy_SBS	reuni 212 memang berkampanye terselubung oleh prabowo
4.	FahrizalHutaini5	tabir semakin terungkapppp banyak peserta 212 adalah hti

Berdasarkan dari contoh tweet table 5.11 yang mendapat opini dari masyarakat tentang isu radikalisme terhadap pemilihan presiden 2019 dari isu radikalisme seperti user Dian_melawati dan Abuy_SBS mengungkapkan di status twitternya sebuah ajakan untuk memilih paslon pilpres 02 yaitu prabowo sandi dengan acara reuni 212 melalui acara tersebut termasuk kampanye terselubung. User Kieraha_Indo mengungkapkan reuni 212 menjadi bukti kalau agama menjadi kepentingan politik untuk memilih prabowo sandi. Kemudian akun FahrizalHutaini5 mengungkapkan bahwa sebagian besar peserta yang ikut dalam acara 212 adalah HTI. Melihat dari kondisi diatas bisa disimpulkan sebagian besar pendukung paslon pilpres 2019 prabowo sandi adalah radikal dan dari radikalisme tersebut bisa memainkan peran penting dalam masyarakat dan mengutamakan agama, dari acara reuni 212 tersebut menyimpulkan sebuah kampanye terselubung untuk

tidak memilih paslon pilpres pasangan tertentu dan hal tersebut sangat berpengaruh terhadap hasil dari pemilihan presiden 2019.

5.4.3 isu hoax

Didalam isu hoax saat pemilihan presiden 2019 menjadi isu yang sangat mempengaruhi hasil tersebut berdasarkan data yang dikumpulkan dari twitter. Terpilihnya isu hoax saat pemilihan presiden 2019 dikarenakan banyak penyebaran berita hoax saat masa pilpres. Diantaranya saat kebanyakan kubu prabowo debat dan dinilai oleh masyarakat langsung dari jejaring sosial twitter dan masyarakat mengungkapkan semuanya di twitter saat prabowo debat. Di dalam table 5.12 tertera beberapa contoh tweet yang menunjukkan isu hoax.

Tabel 5.12 contoh tweet dari isu hoax

No.	Username	Status
1.	Bang_Pino	seorang yg bergaya sederhana, muka lugu apa adanya bisa dipercaya, ternyata itu di media. DIA RAJA TEGA DIA PEMBOHONG doyan utang tapi rakyat yg bayar.. TEGA doyan impor buat petani terkapar.. TEGA Insya Allah #2019PrabowoPresiden
2.	Hendromaz	Lah kan lagi cari sensasi ben populer, biar nggak tenggelam kubu sebelah. #BohongKokPidato Noh...omongan si prabowo ngawurrr....klo mau di naikkan lagi jd berapa? Memperkaya pejabat. Mencekik rakyat jelata dengan pajak yg tinggi.. #bohongkokpidato

3.	CarolinePutri14	Kalau ngga BOHONG & FITNAH bukan pasangan sehoaks @prabowo @sandiuno #prabohongstadium4
4.	AuliaAstuti7	capres dan Cawapres kok ga kompak...inilah kalo kebanyakan bohong. Semua materi debat prabosan BOHONG, mulai kepala desa mojosukerto, nelayan karawan, semua tidak sesuai fakta.

Table 5.12 melihat tanggapan yang beragam dari user pengguna twitter tentang hoax didalam pemilihan presiden 2019 baik itu untuk Jokowi dan prabowo juga. Akun Bang_Pino menyatakan kalau Jokowi sudah berbohong dikarenakan Jokowi yang bergaya sederhana, muka lugu senang untuk menghutang ke negara lain tetapi rakyat yang membayar, dan suka memasukan barang apa saja yang import membuat petani sendiri terkapar. Sedangkan untuk akun Hendromaz, CarolinePutri14, dan AuliaAstuti7 mereka bertanggapan kalau prabowo ngomongnya ngawur, yang diomongkan saat debat pilpres 2019 bisa mencekik masyarakat jelata kalau pajak ditinggikan dan memperkaya pejabat, dan prabowo tidak berbicara tentang fakta, semua mengada ada saat prabowo berbicara tentang visi misi nya dan ditanggapi oleh masyarakat sendiri. Hal tersebut dengan isu hoax sangat menarik saat pasca pemilihan presiden 2019, masyarakat saling lempar tuduhan satu sama lain untuk calon presidennya masing-masing dari banyak nya hoax terhadap paslon pemilihan presiden 2019, dari situ dapat murunkan nilai suatu pasangan dan berpengaruh terhadap hasil pilpres 2019.

5.4.4 isu kampanye

Isu kampanye merupakan factor yang sangat berpengaruh tertinggi juga didalam hasil pemilihan presiden 2019 berdasarkan data twitter yang

berhasil dikumpulkan. Didalam team setiap paslon akan mendukung sepenuhnya calon nya yang didukung, dan mempunyai team sukses, dari team sukses tersebut berperan memainkan peran penting dalam menjalankan kampanye nya untuk memenangkan calon nya masing-masing. Dalam kampanye tersebut setiap paslon mengeluarkan janji-janji dan program mereka untuk memperbaiki Indonesia dan membuat masyarakat percaya kepada mereka agar memilih mereka sebagai presiden. Tetapi team sukses prabowo menggunakan media sosial apapun sebagai media untuk kampanye, dari media sosial tersebut juga mendapatkan pendukung untuk paslon tersebut, dan prabowo juga memanfaatkan aksi reuni 212 untuk kampanye, tetapi berbeda untuk kubu Jokowi, mereka langsung terjun ke masyarakat menjadi media untuk kampanye, strategis langsung kepada masyarakat. Pada table 5.13 terdapat beberapa contoh tweet isu kampanye.

Tabel 5.13 contoh tweet dari isu kampanye

No.	Username	Status
1.	Don_design5	Pak Jokowi dan pak amin selamat ya pak udah jadi presiden Indonesia, selamat menunaikan janji2 kampanye nya
2.	Fahri_ali2	Seorang presiden di beri hak sepenuhnya dlm sebuah keputusan reklamasi. Kita tunggu pak Jokowi tolak reklamasi teluk benoa spt janji kampanye nya
3.	Evryuke	Menagih Janji itu bukan kufur nikmat. Janji=Hutang Masak janji bayar 7% trus dibayar cm 5%, kami tagih yg 2% nya .Trus kami dibilang kufur nikmat? Gini katanya, "Udah

		syukur dibayar 5%... Jangan Kufur Nikmat..." #JKWKufurJanji #jokowi
4.	Bodhy40	Hahaha mmg masih ada yg percaya. ?. Janji kemarin aja di kritik dan di tagih di bilang kufur nikmat. Jgn bodohi rakyat agar mensyukuri sesuatu yg tdk di tepati dan gagal. Rakyat faham mana yg harus di syukuri dan mana yg harus di tagih.

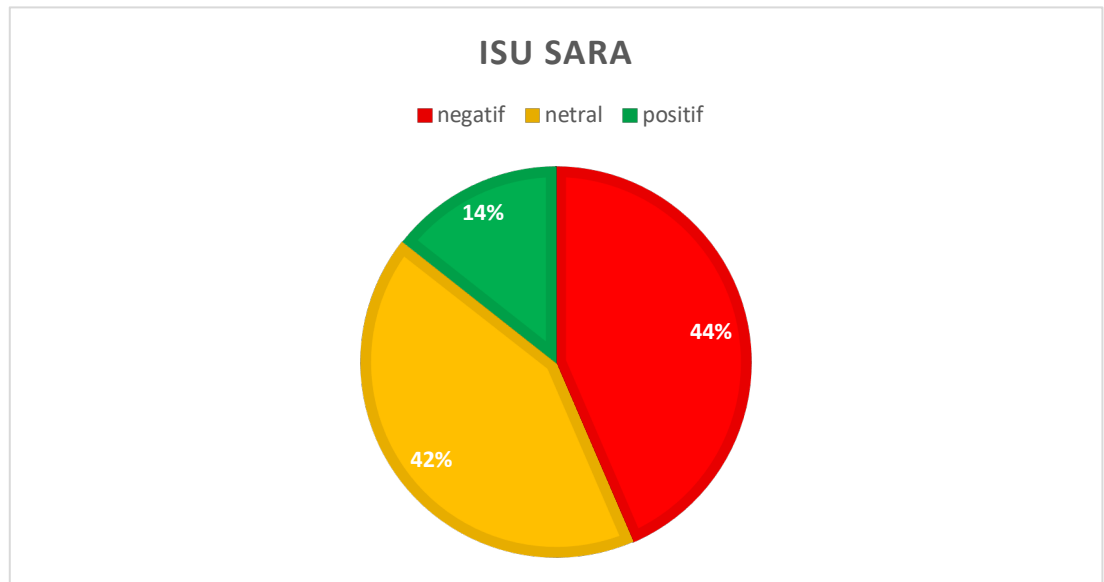
5.5 Analisis sentiment dengan isu yang berpengaruh pada pilpres 2019

Sesudah mendapatkan isu-isu yang dapat mempengaruhi hasil dari Pemilihan Presiden 2019 kemarin, penulis segera melakukan analisa dari isu yang di dapat terhadap sentimennya.

Dari setiap faktor yang mempengaruhi hasil dari Pemilihan Presiden 2019 tersebut dikelompokkan menjadi tiga sentiment kelas, yang didalamnya sentiment itu terdapat sentimen dari negatif, netral, dan positif. Dan ada empat isu yang paling mempengaruhi masing-masing paslon, yaitu ada isu sara, isu radikalisme, isu hoax, dan isu kampanye.

5.5.1 analisis sentiment terhadap isu sara

Jumlah tweet yang didapat dari isu sara sebanyak 39930 tweet, dengan sentiment tertinggi yaitu sentiment negative sebanyak 17383 tweet. Kemudian untuk sentiment netral sebanyak 16791 tweet, dan untuk sentiment positif sebanyak 5756 tweet. Semua hasil yang didapat dalam isu sara tersebut, dapat dilihat pada gambar 5.11 dan table 5.14 memperlihatkan beberapa contoh tweet isu negatif terhadap isu SARA sebagai berikut :



Gambar 5.11 Kelas sentiment isu SARA

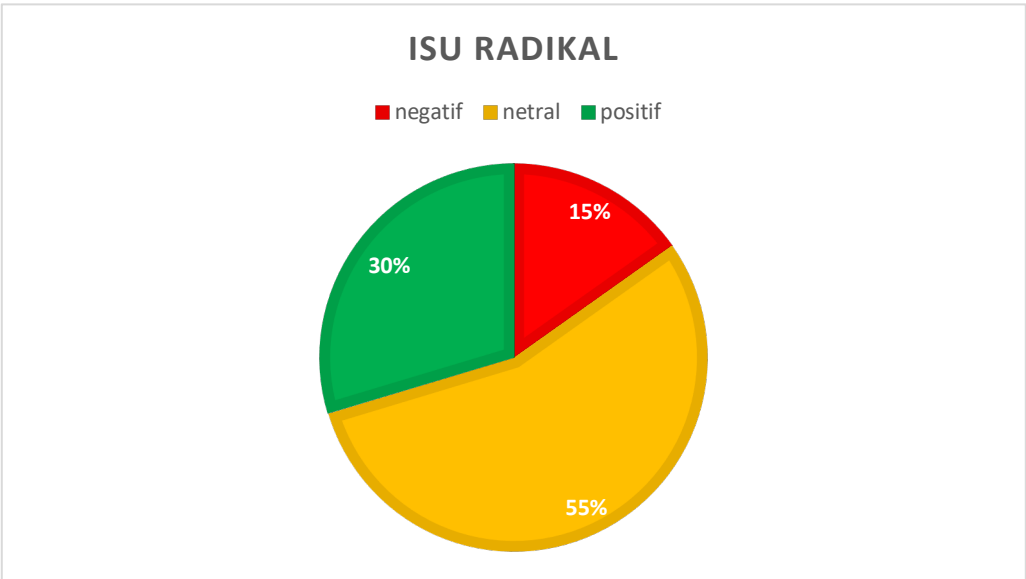
Tabel 5.14 Contoh tweet sentiment negatif terhadap isu SARA

User	Status
heniken110	kalo berbicara bohong [-5] dan menyembunyikan [-2] kebenaran [4] islam gmn dosa [-4] boten jelasislamnya
Aalyahfa	cagub dki janji [-1] 5tahun menjabat eh pilpres nyapres dki dieler2 namanya ambisius nyapres janji [-1] ga import eh jd presiden yg mutus import dustamuabadi [-4] ordesabun 2019gantipresiden [1]
HendryMuhktar	beda yg sdah terbukti [3] berbohong [-5] dg yg terbukti [3] berbohong [-5] makanya hrs pilih prabowo biar tau ordesabun

5.5.2 analisis sentiment terhadap isu radikalisme

Jumlah tweet yang didapat dari isu radikalisme sebanyak 25801 tweet, dengan sentiment tertinggi yaitu sentiment netral 14229 tweet. Kemudian untuk sentiment negatif diperoleh 3919 tweet, dan untuk sentiment positif

diperoleh sebanyak 7653 tweet. Semua hasil yang didapat dalam isu radikal tersebut dapat dilihat pada gambar 5.12 dan untuk tabel contoh sentiment negatif dari isu radikal dapat dilihat di tabel 5.15 sebagai berikut:



Gambar 5.12 Kelas sentiment isu radikal

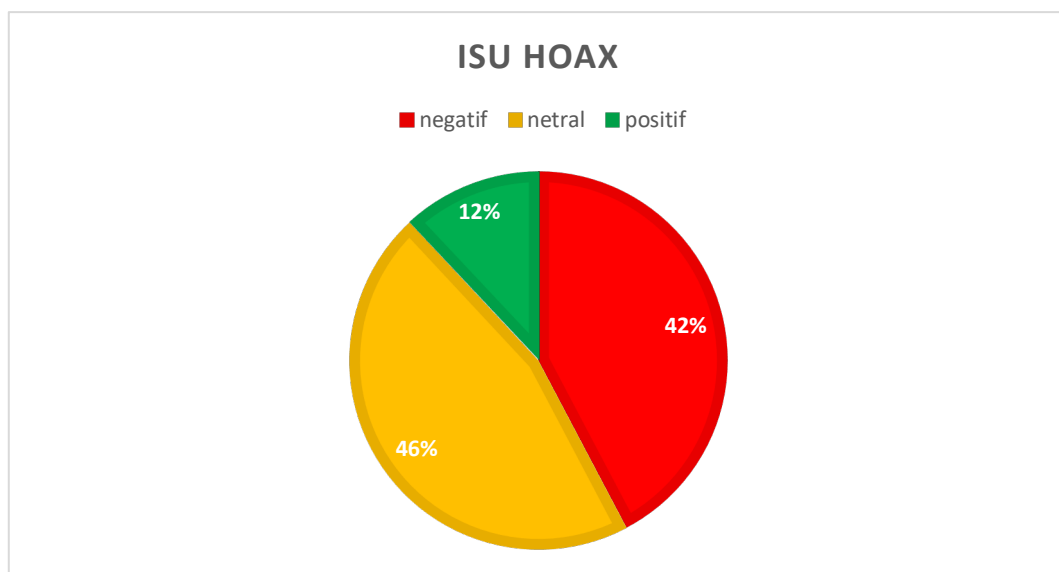
Tabel 5.15 Contoh sentimen negatif terhadap isu radikal

User	Status
Gagas107	wkwkwk bot bot pikir ga misi politik dungu [-4] donk bot hahaha oiya bot ga bisa mikir susah [-3] dipaksa mikirpun ntar yg korslet 2019gantipresiden [1] anakmudapilihprabowosandi papuamemilihprabowosandi sulselpilihprabowosandi
adjatmiko63	pikir rakyat bodoh [-4] dan mudah [3] ditipu [-4] pdhl menjalankan hoax [-4] dustamuabadi [-5] sulselpilihprabowosandi papuamemilihprabowosandi

	prabowosandisolusinkri ahmaddhanikorbanrezim saverockygerung
AJL_Pasee	keadilan jarang [-1] rezim solusi satu2nya tenggelamkan [- 4] rezim ordesabun sambut pemimpin mengangkat martabat negri adilmakmur02 2019prabowopresidenri papuamemilihprabowosandi

5.5.3 analisis sentiment terhadap isu hoax

Jumlah tweet yang didapat dari isu hoax sebanyak 40076 tweet, dengan sentiment tertinggi yaitu sentiment netral 18349 tweet. Kemudian untuk sentiment negatif diperoleh 16948 tweet, dan untuk sentiment positif diperoleh sebanyak 4779 tweet. Semua hasil yang didapat dalam isu radikal tersebut dapat dilihat pada gambar 5.13 dan untuk tabel contoh sentiment negatif dari isu radikal dapat dilihat di tabel 5.16 sebagai berikut:



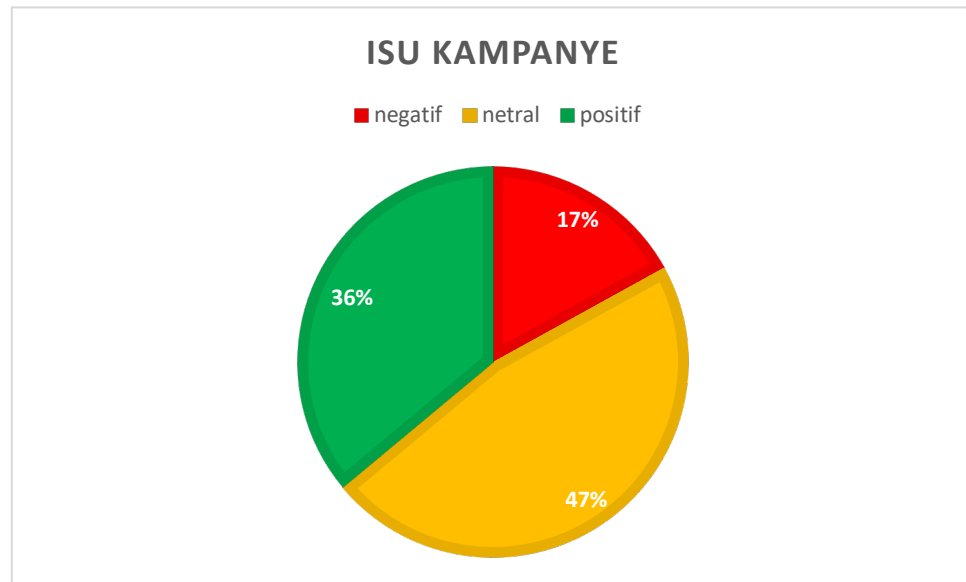
Gambar 5.13 kelas sentimen isu hoax

Tabel 5.16 Contoh sentimen negatif terhadap isu hoax

User	Status
Orang_Atjeh	biar virall [2] donk [2] namanya tukang bohongkokpidato [-3] cari sensasi [-2] mendongkrak popularitas [-1]
gerilyamantika	noh omongan [-2] prabowo ngawur [-4] klo naikkan jd memperkaya pejabat mencekik [-4] rakyat jelata pajak [-3] yg bohongkokpidato [-3]
Ben_salamah777	'ngga bohong [-5] fitnah [-5] pasangan [4] sehoaks prabowo sandiuno prabohongstadium4 [-4] bohongkokpidato [-4] bohongkoknyapres

5.5.4 analisis sentiment terhadap isu kampanye

Jumlah tweet yang didapat dari isu hoax sebanyak 40076 tweet, dengan sentiment tertinggi yaitu sentiment netral 18349 tweet. Kemudian untuk sentiment negatif diperoleh 16948 tweet, dan untuk sentiment positif diperoleh sebanyak 4779 tweet. Semua hasil yang didapat dalam isu radikal tersebut dapat dilihat pada gambar 5.14 dan untuk tabel contoh sentiment negatif dari isu radikal dapat dilihat di tabel 5.17 sebagai berikut:



Gambar 5.14 Kelas sentimen isu kampanye

Tabel 5.17 Contoh sentimen negatif terhadap isu kampanye

User	Status
dyanr99	fitnah [-4] dan hoax [-4] dituduhkan kiyai marufamin kepentingan politik ikutkyai kyaimarufamin jokowiaminmenangdebat [3] santri
Sipendosa8	yg catat sejarah ketololan oposan hoax [-4] oposan dan hattrick kekalahan prabowo subianto jokowiaminmenangdebat [3] jokowilagi
RizkyRaspek	ahelah jokowiaminmenangdebat [3] dimananya menanh tong tonton tuh [-2] yutub visi misi mekanisme anti korupsi [-4] kedepannya kek [-2] tuh [-2] jokowi luh tau yah gw ragu [3] prabowo sempet bilang bikin kpk ri daerah2 kalah [-4] luh

5.6 Analisis hasil dari opini masyarakat

Berdasarkan hasil yang didapat dari sebuah pengelompokan sentimen opini yang dikeluarkan oleh masyarakat terhadap opini yang dapat mempengaruhi hasil dari pemilihan presiden 2019 tersebut didominasi oleh sentimen netral, dan yang kedua didominasi oleh sentimen negatif. Faktor isu yang didominasi oleh sentiment netral adalah isu radikal, isu hoax, dan isu kampanye, sedangkan isu sara didominasi oleh sentimen negatif. Dari kesimpulan diatas dapat disimpulkan bahwa yang paling dominan adalah sentimen netral, dan isu negatif dikarenakan sentimen yang berhasil didapat dari hashtag. Dan sebagian besar factor yang didapat dari sentimen negatif didapat dengan hashtag yang dikelompokkan dengan isu sara.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

6.2 Saran

JADWAL PENELITIAN

	Kegiatan	Agustus 2019				September 2019				Oktober 2019					November 2019				Desember 2019					Januari 2020			
		1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3	4	5	1	2	3	4
1	Menentukan topik dan permasalahan	■	■																								
2	Menentukan lokasi penelitian					■	■																				
3	Menyusun metodologi penelitian					■	■	■	■	■	■																
4	Pengumpulan data						■	■	■	■	■	■															
5	Pengolahan dan analisis data										■	■	■	■	■												
6	Penyusunan laporan skripsi															■	■	■	■	■	■	■					
7	Penyajian laporan skripsi																						■				
8	Ujian pendadaran																							■	■		

DAFTAR PUSTAKA

- Ahmad Fathan Hidayatullah, A. S. (2014, Agustus 12). *ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER*. Retrieved from https://www.academia.edu/9844318/ANALISIS_SENTIMEN_DAN_KLASIFIKASI_KATEGORI_TERHADAP_TOKOH_PUBLIK_PADA_TWITTER?auto=download
- Andre Maureen Pudjajana, D. M. (2018, April 01). *SENTIMEN ANALISIS TWEET PORNOGRAFI KAUM HOMOSEKSUAL*. Retrieved from <http://jurnal.umk.ac.id/index.php/simet/article/viewFile/1922/1258>
- Buntoro, G. A. (2016, Oktober 21). *ANALISIS SENTIMEN HATESPEECH PADA TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE*. Retrieved from https://www.researchgate.net/publication/309322787_ANALISIS_SENTIMEN_HATESPEECH_PADA_TWITTER_DENGAN_METODE_NAIVE_BAYES_CLASSIFIER_DAN_SUPPORT_VECTOR_MACHINE
- Hartanto. (2017, Maret 01). *TEXT MINING DAN SENTIMEN ANALISIS TWITTER PADA GERAKAN LGBT*. Retrieved from <https://journal.unnes.ac.id/nju/index.php/INTUISI/article/view/9561/6187>
- M. Fachrurrozi, M. N. (2016, Juni 14). *ANALISIS SENTIMEN PENGGUNA JEJARING SOSIAL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE*. Retrieved from https://www.researchgate.net/publication/303942719_ANALISIS_SENTIMEN_PENGGUNA_JEJARING_SOSIAL_MENGGUNAKAN_METODE_SUPPORT_VECTOR_MACHINE
- Ira Zulfa, E. W. (2017, Juli 18). *Sentimen Analisis Tweet Berbahasa Indonesia dengan*. Retrieved from <https://jurnal.ugm.ac.id/ijccs/article/viewFile/24716/16691>
- Novita Dewi Susanti, E. S. (2016, Desember 14). *Uji Perbandingan Akurasi Analisis Sentimen*. Retrieved from http://repository.uksw.edu/bitstream/123456789/13368/1/T2_972014004_Full%20text.pdf
- Busman, N. N. (2018, Juni 09). *Analisa Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java untuk Deteksi Kemacetan di Jakarta*. *Jurnal Online Informatika*, 2-9.
- Bertalya. (2009). *Konsep Data Mining*. *Data Mining & Knowledge Discovery in Databases*.
- Buntoro, G. A. (2016). *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter*. *Integer Journal Maret*, 1(1), 32–41.
- Fachrurrozi, M., & Yusliani, N. (2016). *Analisis Sentimen Pengguna jejaring Sosial Menggunakan Metode Support Vector Machine*, (June).
- Haltuf, M. (2011). *UNIVERSITY OF ECONOMICS IN PRAGUE Faculty of Business Administration*.
- Hidayatullah, A. F., & Sn, A. (2014). *Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter*. *Seminar Nasional Informatika 2014*, 2014(August 2013), 0–8.
- Leslie, H., Spits, H., Lumban, F., Trisetyarso, A., & Abdurachman, E. (2017). *Understanding of Data Mining*

in Computer Science, 4–7.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining. AAAI-2011 Tutorial*.

Monarizqa, N., Nugroho, L. E., & Hantono, B. S. (2014). Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro Dan Teknologi Informasi*, 1, 151–155.

Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine, 2(1), 1–7.

Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support Vector Machine. *Encyclopedia of Biometrics*, 1303–1308.

Pak, A., & Paroubek, P. (2016). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Ijarccce*, 5(12), 320–322.

Pang, B., & Lee, L. (2006). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Informatio*Pang, B., & Lee, L. (2006). *Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval*, 1(2), 91–231.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2), 143–157.

Pratama, E. E., & Trilaksono, B. R. (2015). Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM). *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 1(2), 53.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, 7(1), 59–64.

Rossi, A., Lestari, T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1718–1724.

Schonlau, M., Guenther, N., & Sucholutsky, I. (2017). Text mining with n-gram variables. *Stata Journal*, 17(4), 866–881.

Sembiring, K. (2007). Penerapan Teknik Support Vector Machine untuk Pendeteksian Intrusi pada Jaringan. *Penerapan Teknik Support Vector Machine Untuk Pendeteksian Intrusi Pada Jaringan*, (September), 1–28.

Sharma, P., Agrawal, A., Lalit, A., & Garg, A. (2017). Challenges and Techniques in Preprocessing for Twitter Data, 7(4), 6611–6613.

Susilowati, E., Sabariah, M. K., Gozali, A. A., Informatika, J. T., Telkom, U., & Machine, S. V. (2015). Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter Implementation Support Vector Machine Method for Traffic Jam Classification on Twitter, 2(1), 1478–1484.

- Wahid, D. H., & SN, A. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), 207.
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining. Knowledge and Information Systems* (Vol. 14).
- Zainuddin, N., & Selamat, A. (2014). Sentiment analysis using Support Vector Machine. *2014 International Conference on Computer, Communications, and Control Technology (I4CT)*, (I4ct), 333–337.
- Marpaung, F. H. (2017). *Analisis Sentimen Opini Masyarakat Indonesia Mengenai Calon Gubernur DKI Jakarta 2017 Menggunakan Naïve Bayes Classifier*. Yogyakarta.
- AR, M. (2017, Maret 10). *Jenazah Nenek Hindun Ditelantarkan Warga Setelah Pilih Ahok*. Dipetik April 04, 2018, dari www.liputan6.com: <https://www.liputan6.com/news/read/2882270/jenazah-nenek-hindun-ditelantarkan-warga-setelah-pilih-ahok>
- Artharini, I. (2017, April 21). *Pilkada DKI Jakarta: Faktor-faktor apa saja yang menyebabkan kekalahan Ahok-Djarot?* Dipetik April 04, 2018, dari www.bbc.com: <http://www.bbc.com/indonesia/indonesia-39652042>
- Bonasir, R. (2017, Desember 01). *Dengan 212, umat Islam 'ingin berkuasa lewat pilkada, pemilu dan pilpres'*. Dipetik April 04, 2018, dari <http://www.bbc.com>: <http://www.bbc.com/indonesia/indonesia-42184928>
- Faqir, A. A. (2017, Febuari 24). *Anies klaim program rumah DP Rp 0 wujud keadilan buat warga Jakarta*. Dipetik April 04, 2018, dari www.merdeka.com: <https://www.merdeka.com/politik/anies-klaim-program-rumah-dp-rp-0-wujud-keadilan-buat-warga-jakarta.html>
- Fardiansyah, A. (2017, April 03). *Anies-Sandi Sebut Spanduk Bernada SARA Membuat Pilkada DKI 'Tidak Keren'*. Dipetik April 04, 2018, dari www.okezone.com: <https://news.okezone.com/read/2017/04/03/338/1657997/anies-sandi-sebut-spanduk-bernada-sara-membuat-pilkada-dki-tidak-keren>
- Iskandar, R. A. (2017, Januari 17). *Anies Janji Tutup Alexis Jika Terpilih*. Dipetik April 04, 2018, dari www.liputan6.com: <https://www.liputan6.com/pilkada/read/2829174/anies-janji-tutup-alexis-jika-terpilih>
- Khadafi, A. (2017, April 20). *Faktor Agama Menentukan Kemenangan Anies-Sandiaga*. Dipetik April 04, 2018, dari www.tirto.id: <https://tirto.id/faktor-agama-menentukan-kemenangan-anies-sandiaga-cm79>
- Kominfo. (2013, November 07). *Kominfo : Pengguna Internet di Indonesia 63 Juta Orang*. Dipetik April 04, 2018, dari www.kominfo.go.id: https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker
- Kunjana, L. G. (2017, Maret 17). *Tim Pemenangan Anies-Sandi Fokus Kampanye Via Sosmed*. Dipetik April 04, 2018, dari <http://www.beritasatu.com>: <http://www.beritasatu.com/pilkada/420076-tim-pemenangan-anieessandi-fokus-kampanye-via-sosmed.html>

- Pratomo, A. Y. (2017, April 08). *Anies-Sandiaga dianggap bisa diuntungkan adanya isu SARA*. Dipetik April 04, 2018, dari [www.merdeka.com](http://www.merdeka.com/politik/anies-sandiaga-dianggap-bisa-diuntungkan-adanya-isu-sara.html): <https://www.merdeka.com/politik/anies-sandiaga-dianggap-bisa-diuntungkan-adanya-isu-sara.html>
- Ridhoi, M. A. (2017, April 10). *Pakar LIPI: Pilgub DKI 2017 Terburuk dalam Sejarah Pilkada*. Dipetik April 04, 2018, dari www.tirto.id: <https://tirto.id/pakar-lipi-pilgub-dki-2017-terburuk-dalam-sejarah-pilkada-cmqt>
- Rizki, D. (2017, Maret 21). *Alasan Anies-Sandi Tolak Reklamasi karena Tidak Ingin Warga Dimiskinkan*. Dipetik April 04, 2018, dari www.tribunnews.com: <http://wartakota.tribunnews.com/2017/03/21/alasan-anies-sandi-tolak-reklamasi-karena-tidak-ingin-warga-dimiskinkan>
- Safutra, I. (2017, April 18). *1.900 Isu Hoax Muncul Jelang Pencoblosan Pilkada DKI Jakarta*. Dipetik April 04, 2018, dari www.jawapos.com: <https://www.jawapos.com/read/2017/04/18/124206/1900-isu-hoax-muncul-jelang-pencoblosan-pilkada-dki-jakarta>

Lampiran