

**TESIS**

**ANALISIS SENTIMEN PADA TWITTER TERHADAP CALON  
PRESIDEN INDONESIA 2019 MENGGUNAKAN METODE  
NAIVE BAYES, K-NEAREST NEIGHBOR DAN  
PARTICLE SWARM OPTIMIZATION**



Disusun oleh:

**Nama : Abdul Malik Zuhdi**  
**NIM : 17.52.1053**  
**Konsentrasi : Business Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2019**

**TESIS**

**ANALISIS SENTIMEN PADA TWITTER TERHADAP CALON  
PRESIDEN INDONESIA 2019 MENGGUNAKAN METODE  
NAIVE BAYES, K-NEAREST NEIGHBOR DAN  
PARTICLE SWARM OPTIMIZATION**

**TWITTER SENTIMENT ANALYSIS ON INDONESIAN PRESIDENT  
CANDIDATE 2019 USING NAIVE BAYES, K-NEAREST NEIGHBOR  
AND PARTICLE SWARM OPTIMIZATION METHODS**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama** : Abdul Mallk Zuhdi  
**NIM** : 17.52.1053  
**Konsentrasi** : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2019**

**HALAMAN PENGESAHAN**

**ANALISIS SENTIMEN PADA TWITTER TERHADAP CALON PRESIDEN  
INDONESIA 2019 MENGGUNAKAN METODE NAIVE BAYES,  
K-NEAREST NEIGHBOR DAN PARTICLE SWARM  
OPTIMIZATION**

**TWITTER SENTIMENT ANALYSIS ON INDOONESIAN PRESIDENT CANDIDATE  
2019 USING NAIVE BAYES, K-NEAREST NEIGHBOR AND PARTICLE SWARM  
OPTIMIZATION METHODS**

Dipersiapkan dan Disusun oleh

**Abdul Malik Zuhdi**

**17.52.1053**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, 2 Oktober 2019

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer



## HALAMAN PERSETUJUAN

### ANALISIS SENTIMEN PADA TWITTER TERHADAP CALON PRESIDEN INDONESIA 2019 MENGGUNAKAN METODE NAIVE BAYES, K-NEAREST NEIGHBOR DAN PARTICLE SWARM OPTIMIZATION

### TWITTER SENTIMENT ANALYSIS ON INDONESIAN PRESIDENT CANDIDATE 2019 USING NAIVE BAYES, K-NEAREST NEIGHBOR AND PARTICLE SWARM OPTIMIZATION METHODS

Dipersiapkan dan Disusun oleh

**Abdul Malik Zuhdi**

17.52.1053

Telah Dibacakan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari Rabu, 2 Oktober 2019

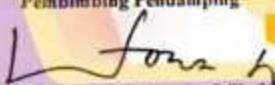
Pembimbing Utama

  
Prof. Dr. Ema Utami, S.Si., M.Kom.  
NIK. 190302837

Anggota Tim Penguji

  
Prof. Dr. Bambang Soediono WA,  
NIK. 55512

Pembimbing Pendamping

  
Dr. Suwanto Rahario, S.Si., M.Kom.  
NIK. 999106


  
Dr. Andi Supvoto, M.Kom  
NIK. 190302052

  
Prof. Dr. Ema Utami, S.Si., M.Kom.  
NIK. 190302039

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 2 Oktober 2019

Direktor Program Pascasarjana

  
Dr. Kusriani, M.Kom.  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Abdul Malik Zuhdi  
NIM : 17.52.1053  
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:  
**Analisis Sentimen pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization**

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.  
Dosen Pembimbing Pendamping : Dr. Suwanto Raharjo, S.Si., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 2 Oktober 2019  
Yang Menyatakan,



Abdul Malik Zuhdi

## HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah, kupersembahkan karyaku untuk orang-orang yang kusayangi :

1. Ayahku (Bapak Kodirin) dan Ibuku (Ibu Khoiriyah), karena bimbingan, motivasi, dukungan serta doa yang tak henti-henti sehingga saya mampu menyelesaikan pasca sarjana ini di umur ke-25.
2. Saudara-saudaraku Mahmudul Ngato', Miftahul Huda, Mambaul Ulum. Terimakasih atas dukungan dan motivasinya.
3. Teman-teman MTI Angkatan 19 B1 yang tidak bisa saya sebutkan satu persatu.
4. TIM RPL yang selalu mensupport dan membantuku dalam kesusahan.
5. TIM Kanopi yang selalu menghiburku tiap buka group dan membuat tesisku lama selesainya.
6. TIM Ngohwah yang selalu mengajak ngopi sehingga membuat tesisku lama selesainya.
7. Untuk dia yang tak perlu saya sebutkan namanya "I LOVE U FULL".

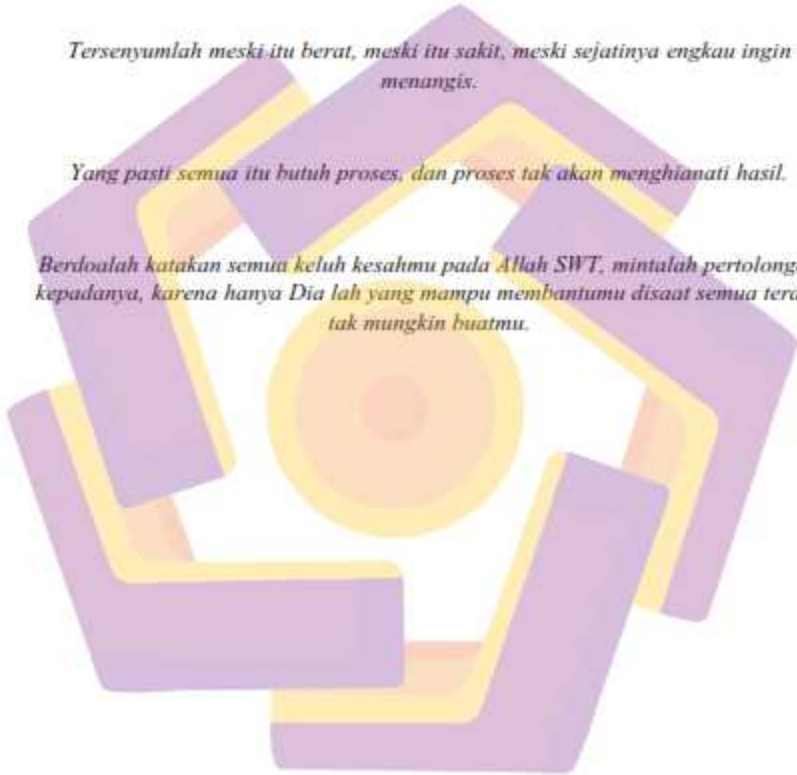
## HALAMAN MOTTO

*Teruslah berjuang dan jangan pernah berhenti meski itu berat dan sangat berat, meski engkau berjuang sendiri, meski banyak orang yang meremehkanmu, dan meskipun engkau tak yakin berhasil.*

*Tersenyumlah meski itu berat, meski itu sakit, meski sejatinya engkau ingin menangis.*

*Yang pasti semua itu butuh proses, dan proses tak akan mengkhianati hasil.*

*Berdoalah katakan semua keluh kesahmu pada Allah SWT, mintalah pertolongan kepadanya, karena hanya Dia lah yang mampu membantumu disaat semua terasa tak mungkin buatmu.*



## KATA PENGANTAR

Dengan Rahmat Allah SWT yang telah memberikan rahmat dan hidayahnya sehingga tesis ini dapat selesai tepat waktu, terima kasih kepada:

1. Prof. Dr. Suyanto selaku rektor AMIKOM, kami bangga bisa menjadi bagian dari AMIKOM.
2. Prof. Dr. Ema Utami, S.Si., M.Kom dan Dr. Suwanto Raharjo, S.Si., M.Kom yang telah dengan sabar membimbing dan memberikan waktu dan pemikirannya hingga selesainya tesis ini.
3. Dr. Kusriani, M.Kom selaku direktur MTI yang telah memberikan kemudahan selama pembelajaran di MTI AMIKOM.
4. Prof. Dr. Bambang Soedijono WA, Dr. Andi Sunyoto, M.Kom dan Prof. Dr. Ema Utami, S.Si., M.Kom selaku Tim Penguji yang telah membimbing dan mengarahkan sampai tesis selesai.
5. Teman teman MTI Amikom Angkatan 19 B1 yang telah berjuang bersama dimanapun kalian berada kita adalah saudara.

Yogyakarta, 2 Oktober 2019

Penulis



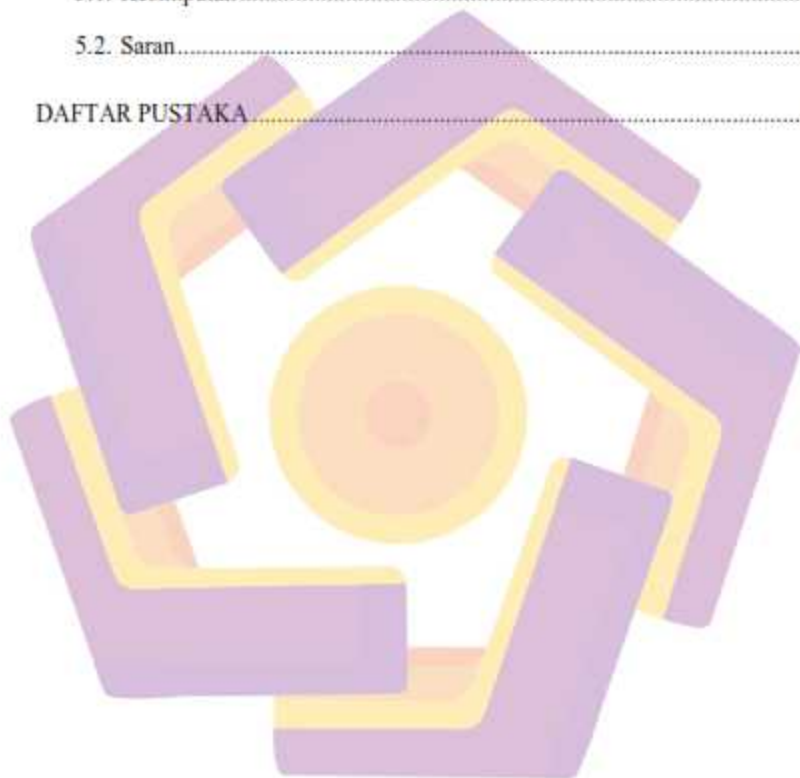
## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERSETUJUAN .....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS .....	v
HALAMAN PERSEMBAHAN .....	vi
HALAMAN MOTTO .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xv
INTISARI .....	xvii
<i>ABSTRACT</i> .....	xviii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah .....	4
1.4. Tujuan Penelitian .....	4
1.5. Manfaat Penelitian .....	4
BAB II TINJAUAN PUSTAKA .....	5

2.1. Tinjauan Pustaka .....	5
2.2. Keaslian Penelitian .....	8
2.3. Landasan Teori.....	17
2.3.1. Analisis Sentimen .....	17
2.3.2. Twitter.....	17
2.3.3. Twitter API .....	19
2.3.4. Text Mining.....	21
2.3.5. Peprocessing.....	22
2.3.6. Opinion Word.....	25
2.3.7. Stemming algoritma dari Nazief-Adriani .....	27
2.3.8. Algoritma klasifikasi .....	28
2.3.9. <i>Naive Bayes</i> .....	28
2.3.10. K-Nearest Neighbor (K-NN).....	31
2.3.11. Particle Swarm Optimization (PSO).....	33
2.3.12. Cross Validation.....	36
<b>BAB III METODE PENELITIAN .....</b>	<b>38</b>
3.1. Jenis, Sifat, dan Pendekatan Penelitian .....	38
3.2. Metode Pengumpulan Data.....	38
3.3. Preprocessing Data.....	40
3.3.1. Transform case.....	41
3.3.2. Cleaning data .....	42

3.3.3. Normalisasi .....	43
3.3.4. Stemming .....	44
3.3.5. Proses penilaian sentiment .....	45
3.4. Klasifikasi sentiment analisis .....	48
3.4.1. Read Excel .....	48
3.4.2. Set role .....	50
3.4.3. Nominal to text .....	50
3.4.4. Process document from data .....	51
3.4.5. Tokenize .....	51
3.4.6. Filter tokens by length .....	52
3.4.7. Filter Stopword Indonesia .....	52
3.4.8. Cross Validation .....	53
3.4.9. Apply model .....	54
3.4.10. Performance .....	55
3.5. Alur penelitian .....	55
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN .....</b>	<b>57</b>
4.1. Crawling data Twitter .....	57
4.2. Seleksi data .....	58
4.3. Preprocessing .....	58
4.4. Sentiment scoring .....	59
4.5. Klasifikasi sentiment analisis .....	60

4.5.2. Klasifikasi sentiment analisis dengan Rapidminer.....	74
4.5.5. Perbandingan klasifikasi sentiment .....	81
<b>BAB V PENUTUP .....</b>	<b>82</b>
5.1. Kesimpulan .....	82
5.2. Saran.....	83
<b>DAFTAR PUSTAKA.....</b>	<b>84</b>



## DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization .....	8
Tabel 2.2 Kumpulan kata positif .....	26
Tabel 2.3 Kumpulan kata Negatif.....	26
Tabel 2.4. Ilustrasi Split Validation.....	37
Tabel 3.1 Transform case .....	41
Tabel 3.2 Cleaning data .....	42
Tabel 3.3 Normalisasi .....	43
Tabel 3.4 Stemming .....	44
Tabel 3.5 sentiment scoring.....	47
Tabel 4.1 Contoh hasil crawling data twitter .....	57
Tabel 4.2 Contoh hasil seleksi data .....	58
Tabel 4.3 Contoh hasil preprocessing data.....	59
Tabel 4.4 Contoh hasil sentiment scoring .....	59
Tabel 4.5 Dataset setelah preprocessing .....	61
Tabel 4.6 Dataset setelah pembobotan TF-IDF.....	61
Tabel 4.7 Perhitungan jarak euclidean.....	62
Tabel 4.8 Perhitungan k tetangga .....	62
Tabel 4.9 Prediksi data testing berdasarkan nilai k .....	63
Tabel 4.10 Perhitungan probabilitas dokumen .....	64

Tabel 4.11 Perhitungan probabilitas kata.....	64
Tabel 4.12 Hasil perhitungan update kecepatan iteasi ke-1 .....	67
Tabel 4.13 Hasil perhitungan update posisi iteasi ke-1 .....	67
Tabel 4.14 Hasil penentuan k tetangga .....	68
Tabel 4.15 Update Pbest iterasi ke-1 .....	68
Tabel 4.16 Jarak euclidean setiap iterasi.....	69
Tabel 4.17 Hasil perhitungan update kecepatan iteasi ke-1 .....	71
Tabel 4.18 Hasil perhitungan update posisi iteasi ke-1 .....	71
Tabel 4.19 Hasil perhitungan probabilitas kata .....	72
Tabel 4.20 Pbest iterasi ke-1 .....	73
Tabel 4.21 nilai probabilitas dokumen baru .....	74
Tabel 4.22 Hasil klasifikasi sentiment jokowi dengan k-NN.....	76
Tabel 4.23 Hasil klasifikasi sentiment prabowo dengan k-NN.....	76
Tabel 4.24 Hasil klasifikasi sentiment jokowi dengan Naive bayes .....	77
Tabel 4.25 Hasil klasifikasi sentiment jokowi dengan Naive bayes .....	77
Tabel 4.26 Hasil klasifikasi sentiment dengan k-NN - PSO.....	79
Tabel 4.27 Hasil klasifikasi sentiment jokowi dengan PSO - NB.....	80
Tabel 4.30 Perbandingan klasifikasi sentiment Jokowi.....	81
Tabel 4.31 Perbandingan klasifikasi sentiment Prabowo .....	81

## DAFTAR GAMBAR

Gambar 3.1 Proses pengumpulan data Twitter.....	39
Gambar 3.2 Parameter Search Twitter.....	39
Gambar 3.3 Hasil Pencarian data Twitter.....	40
Gambar 3.4 <i>Tweet</i> hasil pencarian data.....	40
Gambar 3.5 Preprocessing.....	41
Gambar 3.6 Hasil transform case.....	42
Gambar 3.7 Hasil cleaning data.....	43
Gambar 3.8 Kamus normalisasi.....	43
Gambar 3.9 Hasil Normalisasi.....	44
Gambar 3.10 Hasil Stemming.....	45
Gambar 3.11 Contoh kata positif dan negatif.....	45
Gambar 3.12 Sentiment scoring.....	47
Gambar 3.13 Klasifikasi sentiment analisis.....	48
Gambar 3.14 Tahapan read excel.....	49
Gambar 3.15 Open file.....	49
Gambar 3.16 Tahapan set role.....	50
Gambar 3.17 Tahapan Nominal to text.....	51
Gambar 3.18 Tahapan process document from data.....	51
Gambar 3.19 Tokenize.....	51
Gambar 3.20 Contoh hasil tokenizing.....	52
Gambar 3.21 Tahapan filter tokens by length.....	52
Gambar 3.22 Filer Stopword Indonesia.....	52

Gambar 3.23 Kamus Stopword Indonesia .....	53
Gambar 3.24 Contoh hasil Filter stopwords .....	53
Gambar 3.25 Tahapan cross validation .....	53
Gambar 3.26 Proses di dalam cross validation .....	54
Gambar 3.27 Tahapan apply model .....	54
Gambar 3.28 Tahapan Performance .....	55
Gambar 3.29 Alur Penelitian .....	55
Gambar 4.1 Grafik perbandingan crawling data .....	57
Gambar 4.2 Hasil seleksi data .....	58
Gambar 4.3 Grafik hasil pelabelan .....	60
Gambar 4.4 flowchart KNN .....	61
Gambar 4.5 flowchart Naive bayes .....	63
Gambar 4.6 Penentuan kelas dokumen testing .....	65
Gambar 4.7 flowchart KNN-PSO .....	66
Gambar 4.8 Diagram pergerakan jarak eculidean setiap iterasi .....	69
Gambar 4.9 Flowchart PSO-NB .....	70
Gambar 4.10 Hasil klasifikasi dokumen baru .....	72
Gambar 4.11 Diagram probabilitas kata dokumen baru .....	74
Gambar 4.12 Klasifikasi sentiment analisis .....	75
Gambar 4.13 Proses sentimen analisis menggunakan metode k-NN .....	75
Gambar 4.14 Klasifikasi sentimen dengan metode Naive bayes .....	77
Gambar 4.15 Proses sentimen analisis menggunakan metode k-NN dan PSO .....	78



## INTISARI

Twitter adalah media sosial di mana banyak orang dari seluruh dunia dapat mengungkapkan pendapat mereka. Di Twitter masyarakat ramai berbincang tentang pemilihan pilpres 2019, sehingga bermunculan opini yang bersifat positif maupun negatif tentang calon presiden. Penelitian ini diharapkan dapat bermanfaat membantu riset atas opini masyarakat yang mengandung sentimen positif dan negatif. Tahapan penelitian dilakukan dengan pengumpulan data tweet, preprosesing, tokenisasi, penentuan class atribut, load directory, determine sentiment, kemudian dilanjutkan klasifikasi dengan menggunakan tiga metode yaitu k-NN, Naive Bayes, dan PSO. Dari hasil penelitian dan pengujian, metode K-NN yang di optimasi dengan PSO memiliki tingkat akurasi klasifikasi sentimen mencapai 98,5 % dengan rata-rata waktu 193 detik.

Kata Kunci: Sentimen analisis, Twitter, K-NN, Naive Bayes, PSO.



## **ABSTRACT**

*Twitter is a social media where many people come from all over the world. On Twitter the public was busy talking about the 2019 presidential election, resulting in positive and negative opinions about the presidential candidate. This research is expected to be useful to help community research that contains positive and negative sentiments. Stages of research carried out by collecting tweet data, preprocessing, tokenisasi, class attributes, load directories, determine sentiment, then collected using three methods, namely k-NN, Naive Bayes, and PSO. From the results of research and testing, the K-NN method optimized by PSO has a sentiment classification accuracy rate of 98.5% with an average time of 193 seconds.*

*Keywords: Sentiment analysis, Twitter, K-NN, Naive Bayes, PSO.*



# BAB I PENDAHULUAN

## 1.1. Latar Belakang Masalah

Twitter adalah platform media sosial sebagai tempat yang bagus di mana orang-orang dari seluruh penjuru dunia dapat mengungkapkan pendapat mereka. Twitter menghasilkan sekitar 500 juta tweet setiap hari yang berjumlah sekitar 8 TB data. Data yang dihasilkan di Twitter bisa sangat berguna jika dianalisis karena data tersebut dapat diekstrak menjadi informasi penting melalui opini mining. Pendapat tentang berita atau peluncuran produk apapun atau bahkan jenis tren tertentu dapat diamati dengan baik di Twitter (Dubey, 2017). Berdasarkan data dari [website statistical](#) , perkembangan pengguna Twitter pada tahun 2011 mencapai 117 juta pengguna dan pada tahun 2017 mencapai 328 juta pengguna aktif di seluruh dunia.

Pengguna Twitter di Indonesia pada tahun 2016 menurut [website socialmemos 2](#) mencapai 29 juta pengguna dengan 2,4% dari 10 juta tweet worldwide. Maraknya Pilpres di Indonesia menyebabkan Twitter digunakan untuk kepentingan politik oleh masyarakat atau institusi politik. Twitter berperan aktif dalam proses komunikasi antara institusi politik dengan masyarakat terutama pada saat kampanye berlangsung (Rasyadi, 2017). Selama masa kampanye, banyak berita terbaru tentang capres yang bermunculan, menyebabkan masyarakat beropini positif atau negatif tentang sebuah berita. Pentingnya opini masyarakat terhadap capres menjadi hal utama dalam penelitian ini, karena seorang pemimpin negara

harus disukai oleh semua warga negaranya. Hal ini diyakini menjadikan Twitter sebagai sumber data yang berpotensi dan efisien mengenai Pemilihan Presiden Tahun 2019. Penelitian ini menggunakan Twitter karena memiliki beberapa keuntungan, yaitu digunakan oleh berbagai kalangan pengguna, memiliki pesan singkat disebut tweet yang mengandung opini masyarakat yang beragam, bertambah setiap saat, dan persebaran berita yang cepat.

Pada penelitian sentimen sebelumnya terhadap Twitter @SBYudhoyono dengan menggunakan metode Naive Bayes dihasilkan akurasi 79,42% (Aliandu, 2013). Particle Swarm Optimization (PSO) adalah metode optimasi global yang diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995 berdasarkan penelitian terhadap perilaku kawanan burung dan ikan. Partikel memiliki kecenderungan untuk bergerak menuju daerah pencarian yang lebih baik selama proses pencarian. (Husin Muhamad, 2017). Penelitian oleh Sumanda Ika Novichasari berhasil menunjukkan peningkatan akurasi NBC dengan kombinasi algoritma PSO 4,30% yaitu dari 73,70% menjadi 78,00% pada kasus kredit Bank (Novichasari, 2017).

Mengacu pada penelitian sebelumnya, penelitian ini bertujuan mengetahui sentimen analisis Twitter capres Indonesia tahun 2019. Proses pengumpulan data yang dilakukan pada penelitian ini menggunakan Twitter API pada library Tweepy (Aliandu, 2013). Pengumpulan data dilakukan berdasarkan kata kunci berupa akun capres Indonesia tahun 2019, dengan jumlah tertentu dan pengambilan sampel di tanggal tertentu. Selama pengumpulan data, peneliti mengambil atribut isi teks dan id-user kemudian menambahkan atribut sentimen dan akun. Setelah itu tahap pra-proses, yaitu normalisasi data dengan cara mengubah kata tidak baku menjadi

baku, menghilangkan angka, menghilangkan tanda baca dan simbol, stemming data dengan menghilangkan imbuhan pada setiap kata, dan penghapusan stopword (Husin Muhamad, 2017), kemudian masuk ke tahap sentimen menggunakan metode Naïve Bayes, K-NN dan Particle Swarm Optimization yang diyakini dapat meningkatkan hasil akurasi dari kedua metode di atas dan dihitung jumlah sentimen negatif dan positif.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang di atas maka rumusan permasalahan pada penelitian ini adalah sebagai berikut :

- a. Apakah analisis sentimen tentang opini publik pada capres pilpres Indonesia tahun 2019 dapat diukur menggunakan Naive Bayes, K-NN dan Particle Swarm Optimization?
- b. Berapa akurasi yang didapatkan dari analisis sentimen tentang opini publik pada capres pilpres Indonesia tahun 2019 dengan menggunakan Naive Bayes, K-NN dan Particle Swarm Optimization?
- c. Bagaimana hasil perbandingan analisis sentimen tentang opini publik pada capres pilpres Indonesia tahun 2019 dengan menggunakan Naive Bayes, K-NN dan Particle Swarm Optimization?

### 1.3. Batasan Masalah

Agar penelitian ini terarah dan tidak menyimpang maka diperlukan batasan masalah, diantaranya adalah sebagai berikut:

- Data yang digunakan untuk penelitian ini adalah *tweets* dalam bahasa Indonesia dan *tweets* yang mengandung *keyword* Jokowi dan Prabowo.
- Algoritma yang akan digunakan adalah Naïve Bayes, K-NN dan Particle Swarm Optimization.
- Penelitian ini mengabaikan *tweet* oleh robot, *spam* dan *retweet*.
- Data *tweets* yang digunakan diambil pada bulan April 2019.
- Hasil keluaran pada penelitian ini berupa sentimen positif dan negatif.

### 1.4. Tujuan Penelitian

Tujuan yang akan dicapai melalui penelitian ini yaitu :

- Melakukan analisis sentimen terhadap capres- Indonesia tahun 2019 menggunakan metode Naïve Bayes, K-NN dan Particle Swarm Optimization.
- Mengetahui perbandingan tingkat akurasi Naive Bayes menggunakan Particle Swarm Optimization dan K-NN menggunakan Particle Swarm Optimization.

### 1.5. Manfaat Penelitian

Manfaat penelitian ini adalah :

- Mengetahui sentimen masyarakat terhadap pasangan capres Indonesia 2019.
- Mengetahui perbandingan akurasi antara Naive Bayes, k-NN dan PSO.
- Mengetahui meningkatkan akurasi Naive Bayes dengan PSO dan k-NN dengan PSO.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Tinjauan Pustaka**

Sebelumnya telah banyak dilakukan penelitian tentang analisis sentimen, analisis sentimen kini digunakan untuk kepentingan politik, seperti pada penelitian Hadiyan Rasyadi tentang analisis sentimen pada pemilihan gubernur DKI Jakarta 2017, yang menghasilkan akurasi sebesar 60,60% menggunakan Naive Bayes. Penelitian ini menggunakan sampel data Twitter menggunakan Twitter API dari tanggal 4 sampai 19 April 2017. Data yang diambil menggunakan keyword @AhokDjarot dan @JktMajubersama. Model yang dibuat telah berhasil memprediksi setiap data pada data uji (Rasyadi, 2017). Selain itu pada penelitian yang dilakukan oleh Nuvirta Monarizqa dkk tentang analisis sentimen Twitter sebagai pemberi rating menghasilkan nilai akurasi dari konfigurasi algoritma SVM. Dengan preprocessing yang diterapkan penelitian ini untuk analisis sentimen pada teks berupa tweet berbahasa Indonesia adalah sebesar 73,43%. Konfigurasi ini kemudian menjadi core aplikasi analisis sentimen. Aplikasi analisis sentimen dibangun dengan menggunakan Twitter API untuk mengambil data realtime, pustaka Lucene sebagai alat stemming pada core aplikasi, SVMlight, dan Java. Ketika konfigurasi diterapkan dalam aplikasi analisis sentimen yang dibuat, nilai akurasi ketika menggunakan kata kunci "Jokowi" sebesar 68%, "Prabowo" sebesar 56%, "kalimilk" sebesar 70% dan "sunmor" sebesar 74% (Antonius Rachmat C, 2016).

Algoritma Naive Bayes dikenal dapat digunakan untuk analisis sentimen publik dan dapat memberikan tingkat akurasi yang cukup tinggi, seperti pada penelitian tentang klasifikasi sentimen komentar politik dari Facebook page capres Indonesia tahun 2014 dengan 68 status yang berisi 3400 komentar menghasilkan akurasi 83% menggunakan Naive Bayes. Peningkatan akurasi terjadi karena data netral tidak digunakan dalam sistem (Antonius Rachmat C, 2016). Menurut Rezki Syaputra dan Rachmansyah pada penelitian analisis sentimen pada sosial media Twitter terhadap politik di Indonesia dihasilkan bahwa Naive Bayes dapat digunakan untuk mencari nilai setiap probabilitas dari setiap tweet yang dianalisa. Penarikan data dengan menggunakan media sosial berhasil dilakukan dengan menggunakan Twitter API. Proses crawling data bersifat realtime, data yang ditarik hari ini tidak akan bisa sama pada saat penarikan data esok harinya. Dari hasil uji analisis sentimen pada data Twitter didapat tweet dengan tiga kategori yaitu kategori positif, negatif dan netral. Dengan analisis sentiment dengan menggunakan text mining yang didalamnya terdapat tahapan preprocessing dan ekstraksi fitur untuk menghasilkan tweet kotor menjadi tweet bersih (Syaputra, 2016).

Particle Swarm Optimization (PSO) adalah algoritma yang dipercaya mampu meningkatkan akurasi dari Naive Bayes. Seperti pada penelitian Husin Ahmad, dkk tentang optimasi Naive Bayes classifier dengan menggunakan PSO pada data iris didapatkan rata-rata fitness tinggi yaitu 97,39 pada percobaan ke 3. Dengan kombinasi tersebut berhasil meningkatkan akurasi Naive Bayes (Husin Muhamad, 2017). Pada penelitian lain Sumanda Ika Novichasari berhasil menunjukkan peningkatan akurasi Naive Bayes Classifier dengan kombinasi



algoritma PSO 4,30% yaitu dari 73,70% menjadi 78,00% pada kasus kredit Bank (Novichasari, 2017). Penelitian (Taufik Hidayatulloh, 2016) yang berjudul Klasifikasi Sel Tunggal PAP Smear Berdasarkan Analisis Fitur Berbasis Naive Bayes Classifier dan Particle Swarm Optimization menunjukkan peningkatan akurasi dari penelitian awal yaitu 62,67%, setelah mengkombinasikan NBC dan PSO akurasi menjadi 95,70%.

K-NN adalah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang menggunakan jarak terdekat atau kemiripan pada objek tertentu. Menurut (Deviyanto, 2018) pada penelitiannya tentang analisis sentimen pada pengguna Twitter menggunakan k-NN menghasilkan akurasi 67,2%. Menurut (Manongga, 2018) pada penelitian sentimen analisis tweet pornografi kaum homoseksual Indonesia di Twitter mendapatkan perbandingan hasil akurasi antara Naive Bayes dan k-NN yaitu Naive Bayes sebesar 87,48% dan k-NN 85,40%.

## 2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1.	Analisis Twitter untuk mengetahui karakter seseorang menggunakan algoritma <i>Naive Bayes</i> classifier (Mohammad Zoqi Sarwani, 2015)	Mohammad Zoqi Sarwani, Wayan Firdaus Mahmudy, Seminar Nasional Sistem Informasi Indonesia, 2015	Analisis kepribadian seseorang pada perusahaan dapat dijadikan pertimbangan dalam perekrutan karyawan maupun untuk kenaikan jabatan. Analisis kepribadian secara konvensional membutuhkan beberapa sumber daya seperti ruangan dan waktu. Penelitian ini memberikan solusi dengan cukup menggunakan media Twitter berdasarkan hasil <i>tweets</i> dari seseorang.	<ol style="list-style-type: none"> <li>1. Dalam penelitian ini dilakukan analisis terhadap <i>tweets</i> atau posting yang terdapat pada Twitter untuk mengetahui karakter kepribadian seseorang.</li> <li>2. Hasil penelitian membuktikan bahwa Twitter dapat digunakan sebagai salah satu media untuk mengetahui kepribadian seseorang melalui posting atau <i>tweets</i> mereka. Selain itu, proses pengklasifikasian Twitter menggunakan metode <i>Naive Bayes Classification</i> juga mampu memberikan tingkat akurasi yang baik dengan membandingkan hasil klasifikasi dari sistem dengan hasil dari pakar.</li> </ol>	Untuk pengembangan dari penelitian ini, perlu untuk melakukan percobaan dengan menggunakan jumlah data latih dan data uji yang besar untuk menghitung keakurasian metode yang digunakan dalam penelitian ini.	Pada penelitian ini hanya menggunakan satu algoritma yaitu <i>Naive Bayes</i> . Penelitian yang akan dilakukan akan menggunakan dua metode yaitu <i>Naive Bayes</i> , <i>k-NN</i> dan <i>PSO</i> .

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2.	Klasifikasi <i>Tweets</i> pada Twitter dengan menggunakan metode fuzzy K-Nearest Neighbour (Fuzzy K-NN) dan Query Expansion Berbasis Apriori (Joda Pahlawan Romadhona Tanjung, 2017)	Joda Pahlawan Romadhona Tanjung, Mochammad Ali Fauzi, Indriati, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2017	Ketika seorang pengguna mencari sebuah berita atau informasi yang diinginkan, permasalahan yang muncul adalah pengguna menjadi kesulitan untuk memilahnya. Proses klasifikasi dapat dilakukan untuk mengkategorikan sebuah <i>tweets</i> dengan menggunakan algoritme Fuzzy KNearest Neighbour. Namun, proses pengklasifikasian sebuah <i>tweets</i> sukar dilakukan karena <i>tweets</i> berupa short-text.	<ol style="list-style-type: none"> <li>1. Pada penelitian yang dilakukan menghasilkan akurasi terbaik sebesar 82%.</li> <li>3. Akurasi terbaik didapatkan saat menggunakan metode Fuzzy K-NN dengan Query Expansion tanpa preprocessing serta threshold untuk nilai support <math>\geq 0.15</math> dan nilai confidence <math>\geq 1</math>.</li> </ol>	<ol style="list-style-type: none"> <li>1. Proses ekspansi dengan algoritme query expansion membutuhkan waktu yang lama saat proses running. Oleh karena, itu perlu dilakukan optimasi algoritme agar proses running data tidak memakan waktu yang sangat lama.</li> <li>2. Diperlukan pengujian klasifikasi dengan menggunakan algoritme klasifikasi selain</li> <li>3. Fuzzy K-Nearest Neighbour atau <i>Naive Bayes</i> untuk mengetahui apakah algoritme query expansion juga dapat diterapkan untuk metode klasifikasi</li> </ol>	Pada penelitian ini hanya menggunakan satu algoritma yaitu <i>Naive Bayes</i> , Penelitian yang akan dilakukan akan menggunakan metode <i>Naive Bayes</i> , k-NN dan PSO.

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					<p>lainnya serta untuk mengetahui metode klasifikasi yang memiliki nilai akurasi terbaik jika digabungkan dengan query expansion.</p> <p>Text preprocessing yang dilakukan peneliti belum sepenuhnya sempurna karena masih belum bisa menangani masalah seperti penghapusan awalan dan akhiran secara efektif pada tahap stemming. Oleh karena itu, pada penelitian selanjutnya diperlukan algoritme stemming yang lebih baik lagi untuk hasil klasifikasi yang lebih akurat.</p>	

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3.	Analisis Sentimen pada Twitter mahasiswa menggunakan metode backpagation (Robert Habibi, 2016)	Robert habibi, djoko budyanto setyohadi, ernawati, INFORMATIKA, 2016	Pada penelitian ini, bagaimana mengetahui kecenderungan emosi mahasiswa dengan analisis sentimen pada Twitter menggunakan metode backpropagation. Hasil analisis sentimen merupakan kecenderungan emosi mahasiswa. Kecenderungan emosi mahasiswa dapat digunakan sebagai acuan untuk menentukan perlakuan yang sesuai terhadap mahasiswa pada saat proses belajar	Sistem analisis sentimen pada Twitter mahasiswa menggunakan metode backpropagation dapat mengenali sentimen pada Twitter mahasiswa. Sentimen yang terkandung merupakan representasi emosi dan kecenderungan emosi mahasiswa. Hasil pengenalan kecenderungan 1.1.1. emosi mahasiswa digunakan sebagai acuan untuk memberikan perlakuan yang sesuai sehingga dapat meningkatkan keberhasilan dalam proses belajar mahasiswa.	i. Penelitian hanya menggunakan algoritma backpropagation.	Pada penelitian ini hanya menggunakan satu algoritma yaitu backpropagation. Penelitian yang akan dilakukan akan menggunakan metode yaitu <i>Naive Bayes</i> , k-NN dan PSO.
4.	Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes	Antonius Rachmat C, Yuan Lukito, JUISI, 2016.	penggunaan metode <i>Naive Bayes</i> untuk melakukan klasifikasi sentimen positif atau negatif terhadap komentar dari status kampanye politik dari Facebook Page.	1. Penelitian ini telah mampu mengembangkan sistem klasifikasi sentimen berdasarkan data Pemilu Presiden Indonesia 2014 dari Facebook Page menggunakan <i>Naive Bayes</i> .	ii. Penelitian menggunakan Facebook	Penelitian yang akan dilaksanakan menggunakan Twitter dan metode yang digunakan adalah <i>Naive</i> .

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	(Antonius Rachmat C, 2016)			<p>2. Dalam implementasinya, sistem klasifikasi yang berasal dari data media sosial berbahasa Indonesia membutuhkan preprocessing terutama dalam hal konversi singkatan dan emoticon.</p> <p>3. Algoritma <i>Naive Bayes</i> mampu mengklasifikasikan sentimen dengan tingkat akurasi rata-rata tertinggi 82%.</p> <p>Peningkatan hasil akurasi terjadi karena data netral tidak digunakan dalam sistem.</p>		<i>Bayes</i> , k-NN dan PSO.
5.	Opinion Extraction of Public Figure Based on Sentiment Analysis in Twitter (Nur Hayatin, 2014)	Nur Hayatin, Mustika Mentari, and Abdilatul Izzah, IPTEK Journal of Engineering, 2014	Tujuan penelitian ini adalah untuk mengekstrak pendapat tentang publik figur berdasarkan analisis sentimen menggunakan specific term dan jumlah followers.	Berdasarkan 6 figur publik yang di amati menggunakan NBC menghasilkan bahwa semua tokoh masyarakat mendapatkan respon yang baik. Hasil yang didapat precision positif sebesar 99% dan negatif sebesar 28%, recall positif 75% dan negatif 95%, dan accuracy 76,67%. Akhirnya,	ii. Dengan menggunakan metode NBC menghasilkan akurasi sebesar 76,67%, mungkin akan lebih tinggi lagi akurasi apabila menggunakan metode <i>Naive Bayes</i> .	Akurasi yang dihasilkan dengan NBC hanya 76,67%. Pada penelitian yang akan dilakukan menggunakan metode -

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				enam tokoh masyarakat mendapat nilai probabilitas di atas 0,5 untuk sentimen positif. Itu artinya mereka semua mendapat a. respon positif dari masyarakat.		metode yang diyakini akan menghasilkan akurasi lebih tinggi.
6.	Clustering and Sentiment Analysis on Twitter Data (Dubey, 2017)	Shreya Ahuja and Gaurav Dubey, TEL-NET, 2017	Tujuan utama penelitian ini adalah untuk menemukan emosi, opini, subyektivitas dan sikap dari natural text dan menghasilkan opini positif dan negatif.	Penelitian ini dapat mengelompokkan skor sentimen dan menentukan rasionya dari opini positif dan negatif yang nyata, bukan sentimen dari sebuah kalimat.	1. Data yang di pilih hanya khusus untuk IPL 2017 v. Tidak menampilkan dengan pasti akurasi dari penelitian yang dikerjakan.	Kekurangan data akan menghasilkan tingkat akurasi yang rendah, untuk itu demi meningkatkan akurasi maka data yang digunakan akan lebih banyak.
7.	Optimasi <i>Naive Bayes</i> classifier dengan	Husin Muhamad , Cahyo Adi Prasajo, Nur Afifah Sugianto	Probabilitas pada <i>Naive Bayes</i> Classifier tidak bisa mengukur seberapa besar tingkat keakuratan sebuah prediksi	Pada optimasi <i>Naive Bayes</i> Classifier dengan menggunakan <i>Particle Swarm Optimization</i>	1)	Penelitian ini digunakan untuk dasar bahwa PSO

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	menggunakan <i>Particle Swarm Optimization</i> pada data iris (Novichasari, Particle Swarm Optimization Meningkatkan Akurasi Naive Bayes Classifier, 2017)	. Listiya Surtiningsih, Imam Cholissodin, JTIK, 2017	selain itu metode <i>Naive Bayes</i> juga memiliki kelemahan pada seleksi atribut. Untuk menyelesaikan permasalahan tersebut, algoritma <i>Particle Swarm Optimization</i> (PSO) dapat digunakan untuk melakukan pembobotan atribut untuk meningkatkan akurasi <i>Naive Bayes</i> classifier.	pada data iris. Klasifikasi dilakukan dengan menentukan bobot atribut optimum dengan menggunakan <i>Particle Swarm Optimization</i> . Hasil klasifikasi diperoleh dari fitness tertinggi.		dapat dikombinasikan dengan <i>Naive Bayes</i> dan k-NN untuk meningkatkan akurasi.
8.	Klasifikasi Sel Tunggal PAP Smear Berdasarkan Analisis Fitur Berbasis <i>Naive Bayes</i> Classifier dan <i>Particle Swarm Optimization</i>	Taufik Hidayatulloh, Asti Herliana, Toni Arifin, SWABUMI, 2016	Mengkombinasikan metode NBC dengan PSO untuk mendapatkan hasil akurasi yang baik di penelitian tentang kanker serviks.	1. Hasil dari analisa dan percobaan yang sudah dilakukan untuk klasifikasi kelas normal dan abnormal yang paling tinggi didapatkan dari data training 826 data (90%) dan testing 92 (10%). Klasifikasi kelas 1 sampai dengan kelas 7 akurasi yang paling tinggi didapatkan dari data	1. Menerapkan metode lain untuk klasifikasi, yaitu dengan menggunakan metode Support vectore machine, neural network, Genetic algorithm dan lain-lain. 2) Menerapkan algoritma terpilih	Penelitian yang akan dilakukan menggunakan metode yang sama namun pada kasus yang berbeda serta menambahkan metode k-NN



Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	(Taufik Hidayatulloh, 2016)			training 734 (80%) dan testing 183 (20%).	dengan membuat aplikasi untuk memudahkan ahli patologi dalam pengamatan sel Pap Smear.	sebagai perbandingan.
9.	Sentimen Analisis <i>Tweet</i> Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan <i>Naive Bayes</i> (Manongga, 2018)	Andre Maureen Pudjajana, Danny Manongga, SIMETRIS, 2018	Mengetahui perbandingan tingkat akurasi antara k-NN dan <i>Naive Bayes</i> dalam sentimen analisis di Twitter tentang kaum homoseksual.	1. Hasil perbandingan kedua metode tersebut berdasarkan total rata-rata pengujian kelima hashtag adalah <i>Naive Bayes</i> sebesar 87,48%, dan k-NN sebesar 85,40%. a. Jadi, metode <i>Naive Bayes</i> memiliki nilai akurasi lebih baik dari metode k-NN.	a. Penelitian ini menggunakan metode <i>Naive Bayes</i> dalam melakukan klasifikasi dan untuk mengetahui keakurasian dari metode tersebut maka dilakukan perbandingan dengan k-NN.	Penelitian yang akan dilakukan menggunakan PSO untuk menambahkan akurasi dari <i>Naive Bayes</i> dan k-NN sehingga akurasi lebih banyak.
10.	Analisis Sentimen Twitter Menggunakan Metode K-Nearest	Defri Rosdiansyah, Surya Agustian, UIN SUKSA Riau, 2014	penelitian ini adalah menentukan analisis sentimen dari kasus politik, khususnya sentimen publik terhadap tokoh politik pada	1. Sistem telah dapat menentukan orientasi sentimen pada media Twitter dengan tingkat akurasi k-NN adalah 70 %, Lexicon 77 % ,	b. Masih terdapat kesalahan orientasi sentiment pada kalimat sarkasme.	Penelitian yang akan dilakukan akan menggunakan metode <i>Naive</i>

Tabel 2.1. Matriks literatur review dan posisi penelitian Analisis Sentimen Pada Twitter Terhadap Calon Presiden Indonesia 2019 Menggunakan Metode Naive Bayes, K-Nearest Neighbor dan Particle Swarm Optimization (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Neighbor dan Pendekatan Lexicon (Defri Rosdiyansyah, 2015)		kandidat calon presiden Republik Indonesia 2014.	<p>Kombinasi A 78 % dan Kombinasi B 82 %</p> <ol style="list-style-type: none"> <li>2. Sistem telah mampu melakukan identifikasi kalimat negasi dan kalimat netral dengan</li> <li>3. menggunakan pendekatan Lexicon.</li> <li>4. Adanya kombinasi metode antara algoritma k-NN dan pendekatan Lexicon berdasarkan pengujian terbukti meningkatkan akurasi</li> <li>5. prediksi orientasi sentimen Twitter.</li> </ol> <p>1.1. Identifikasi klausa Tapi berdasarkan analisa yang telah ada masih belum maksimal hal ini dikarenakan terdapat kalimat yang tidak saling bertentangan sebelum klausa Tapi dan sesudah klausa Tapi.</p>		Bayes, k-NN dan PSO. Diharapkan dengan penggunaan metode tersebut dapat menghasilkan penelitian yang lebih bagus.

## **2.3. Landasan Teori**

### **2.3.1. Analisis Sentimen**

Analisis sentimen juga dikenal sebagai opinion mining atau emotion artificial intelligence adalah penggunaan pemrosesan bahasa alami, analisis teks, komputasi linguistik dan biometrik untuk mengidentifikasi, mengekstrak, menghitung dan mempelajari informasi subjektif secara sistematis (Rizal, 2017). Analisis Sentimen atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan text mining yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Defri Rosdiyansyah, 2015).

Analisis sentimen adalah sebuah teknik untuk mendeteksi opini terhadap suatu subyek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data (Nuswaka, 2003). Analisis sentimen digunakan untuk memahami komentar yang diciptakan oleh pengguna (internet) dan menjelaskan bagaimana sebuah produk maupun brand diterima oleh mereka. (Cvijikj, 2013). Dari beberapa pendapat di atas, bisa diambil kesimpulan bahwa analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif.

### **2.3.2. Twitter**

Twitter adalah situs web dimiliki dan dioperasikan oleh Twitter, Inc., yang menawarkan jaringan sosial berupa *microblog*. Disebut *microblog* karena situs ini memungkinkan penggunanya mengirim dan membaca pesan blog seperti pada

umumnya namun terbatas hanya sejumlah 140 karakter yang ditampilkan pada halaman profil pengguna.. Twitter memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus.Pesan dalam Twitter dikenal dengan sebutan *tweet* (Aliandu, 2013).

Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di internet, dan dijuluki dengan "pesan singkat dari Internet (Rizal, 2017).

Twitter diciptakan untuk menjadi tempat saling berbagi pengalaman antar sesama penggunanya tanpa adanya sekat penghalang. Dengan menggunakannya, pengguna akan mudah untuk mengikuti tren, cerita, informasi dan berita dari seluruh penjuru dunia. Twitter juga membantu penggunanya untuk selalu terhubung dengan orang-orang terdekatnya. Ketika penggunanya mengirimkan *tweet*, pesan tersebut bersifat publik dan bisa diakses oleh siapapun, dimanapun dan kapanpun. 16 Bahkan, bagi orang-orang yang mengikuti (*follow*) akun Twitter tersebut, *tweet* tersebut akan secara otomatis muncul di lini masa orang tersebut. Berikut ini adalah beberapa istilah yang dikenal dalam Twitter:

#### A. *Mention*

*Mention* adalah menyebut atau memanggil pengguna Twitter lain dalam sebuah *tweet*. Mention dilakukan dengan menuliskan '@' diikuti dengan nama pengguna lain.

## B. *Hashtag*

*Hashtag* digunakan untuk menandai sebuah topik pembicaraan di Twitter. Penulisan *hashtag* dimulai dengan tanda '#' diikuti dengan topik yang sedang dibahas. *Hashtag* biasa digunakan untuk meningkatkan visibilitas *tweet* pengguna.

## C. *Emoticon*

*Emoticon* adalah ekspresi wajah yang direpresentasikan dengan kombinasi antara huruf, tanda baca dan angka. Pengguna biasa menggunakan *emoticon* untuk mengekspresikan mood yang sedang mereka rasakan.

## D. *Trending Topics*

Jika *hashtag* adalah cara untuk menandai sebuah topik pembicaraan di Twitter, maka *trending topics* adalah kumpulan dari topik pembicaraan yang sangat populer di Twitter.

Sebuah fakta yang cukup mengejutkan disampaikan oleh Giummole bahwa *trending topics* di Twitter akan meningkatkan prediksi hasil pencarian di Google. Hal ini bisa terjadi karena dalam setiap lima menit, Twitter akan mengeluarkan daftar topik yang sangat populer (*trending topics*) dengan cara memonitor dan menganalisa pembicaraan, sedangkan Google juga akan mengeluarkan daftar pencarian populer yang dicari oleh penggunanya pada setiap jam (Giummole, 2018).

### 2.3.3. Twitter API

API (*Application Programming Interface*) merupakan sekumpulan perintah, fungsi, dan protokol yang dapat digunakan dalam membangun perangkat lunak untuk sistem operasi tertentu, juga merupakan suatu dokumentasi yang terdiri

dari antar muka, fungsi, kelas, struktur untuk membangun *software*. Seorang programmer akan lebih mudah dalam membongkar suatu *software* untuk kemudian dapat dikembangkan atau diintegrasikan dengan perangkat lunak yang lain melalui API. Dengan demikian API dapat dikatakan sebagai penghubung antar aplikasi yang satu dengan aplikasi yang lainnya. Suatu rutin standart yang memungkinkan *developer* menggunakan *system function* dimana *operation system* berperan dalam mengelola hal ini. Sehingga API ini mempunyai keunggulan dalam hal interaksi antar aplikasi.

Keuntungan menggunakan API adalah sebagai berikut :

- a. Portabilitas. *Developer* yang menggunakan API dapat menjalankan programnya dalam sistem operasi apapun asalkan sudah ter-*install* API tersebut.
- b. Lebih mudah dimengerti. API menggunakan bahasa yang lebih terstruktur dan mudah di mengerti daripada bahasa *system call*. Hal ini sangat penting dalam hal *editing* dan pengembangan.

Twitter API yaitu sebuah aplikasi yang diciptakan oleh pihak Twitter agar mempermudah pihak *developer* lain untuk mengakses informasi web Twitter tersebut dengan ketentuan dan syarat yang berlaku seperti yang terdapat pada <http://dev.Twitter.com/oauth>. Twitter API menyediakan beberapa fungsi untuk melakukan suatu tugas tertentu, sehingga pengembang perangkat lunak hanya memanggil fungsi tersebut di dalam perangkat lunak yang dibangun.

Ada beberapa jenis Twitter API :

a. Twitter REST API

Terdiri dari Twitter REST dan Twitter Search. Twitter REST memberikan *core* data dan *core* Twitter *objects*. Twitter search berfungsi untuk melakukan pencarian mengenai suatu *instance* objek Twitter maupun mencari *trend*.

b. Twitter Streaming API

API ini biasa digunakan untuk penggalian data karena melalui API ini informasi bisa didapatkan secara *realtime* dengan volume yang sangat tinggi.

#### 2.3.4. Text Mining

Text mining adalah bidang interdisipliner yang mengacu pada pencarian informasi, pertambangan data, pembelajaran mesin, statistik, dan komputasi *linguistik*. Dikarenakan kebanyakan informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, text mining diyakini memiliki potensi nilai komersial tinggi. Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Syaputra, 2016).

Text mining merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan alat analisis tertentu. Text mining dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk *summarization*, pencarian dokumen teks dan *sentiment analysis*. Text mining bertujuan untuk mencari kata-kata yang dapat mewakili apa

yang ada didalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen.

Text mining mempunyai 5 tahapan yaitu *Tokenizing*, *Filtering*, *Stemming*, *Tagging*, dan *Analyzing* yang dapat dijelaskan dibawah ini :

- a. Tahapan *Tokenizing* adalah proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen.
- b. Tahapan *Filtering* adalah suatu proses dimana diambil sebagian data dari data tertentu, dan membuang data pada frekuensi yang lain.
- c. Tahapan *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya.
- d. Tahapan *Tagging* adalah kata yang belum lama dilahirkan. Dahulu sebelum ada tagging, dunia informasi yang ada di internet berserakan dan tidak tersusun berdasarkan kategorinya. Hal itu bagaikan, perpustakaan tanpa ada pengurusnya atau pustakawan.
- e. Tahapan *Analyzing* yaitu untuk mencari seberapa jauh keterhubungan antar kata-kata setiap dokumen.

### 2.3.5. Preprocessing

Data *tweet* yang telah diambil dari Twitter masih berupa data mentah maka dari itu dilakukan tahap *preprocessing* untuk mendapatkan data bersih agar dapat diproses ke tahap selanjutnya. Tahapan yang dilakukan adalah *cleansing data*, *case folding*, *tokenizing* dan *filtering* yang terdiri dari *stopword removal* serta *stemming*.



a. *Cleansing Data*

Dilakukan untuk mengurangi noise pada data *tweet*. Kata-kata yang tidak penting dihilangkan seperti URL, *hashtag* (#), *username* (@username), email, *emoticon* (:@, :\*, :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya. Contoh :

Input : @LazadaIDCare kecewa dengan pelayanan dari CS Lazada, kemarin saya diinfokan bisa loh utk aktivasi kembali akun.

Output : kecewa dengan pelayanan dari CS Lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun

b. *Case Folding*

*Casefolding* merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai dengan z yang diterima. Karakter selain huruf dihilangkan dan dianggap pembatas (Syaputra, 2016). Dalam penulisan *tweet*, pasti terdapat perbedaan bentuk huruf, tahapan ini merupakan proses merubah bentuk huruf menjadi huruf kecil (*lower case*) atau dapat disebut juga penyeragaman bentuk huruf. Contoh :

Input : kecewa dengan pelayanan dari CS Lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun

Output : kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun

c. *Tokenizing*

Proses *tokenizing* atau *parsing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses *tokenizing* adalah pemenggalan kalimat menjadi kata. Contoh :

Input : kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan  
bisa loh utk aktivasi kembali akun

Output : kecewa, dengan, pelayanan, dari, cs, lazada, kemarin, saya,  
diinfokan, bisa, loh, untuk, aktivasi, kembali, akun

d. *Stopword Removal*

*Stopword* adalah proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen. Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena memberikan pengaruh yang tidak baik dalam proses *text mining* seperti kata-kata "dan", "saya", "kamu", "dengan", "dia" dan lain-lain (Syaputra, 2016). *Stopword Removal* merupakan proses menghilangkan daftar kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan seperti "yang", "di", "ke", "itu" dan lain sebagainya. Contoh :

Input : kecewa dengan pelayanan dari cs lazada kemarin saya,  
diinfokan, bias, loh, untuk, aktivasi, kembali, akun

Output : kecewa pelayanan cs lazada kemarin saya info bisa aktivasi  
kembali akun

e. *Stemming*

*Stemming* adalah salah satu proses dari *Information Retrieval* (IR) untuk mencari kata dasar (Sismono, 2005). *Stemming* adalah tahapan untuk membuat kata berimbuhan menjadi kata dasar sesuai dengan aturan bahasa Indonesia yang benar.

Contoh :

Input : kecewa pelayanan cs lazada kemarin saya bisa aktivasi kembali akun

Output : kecewa pelayanan lazada kemarin saya bisa aktif kembali akun

f. *Convert Negation*

Dalam Bahasa Indonesia terdapat kata “tidak”, “nggak”, “tak”, “kurang”, “tanpa” yang disebut kata negasi yaitu kata yang dapat membalikan arti dari kata yang sebenarnya. Contoh :

Input : nggak suka belanja di lazada kualitas barang jelek

Output : nggak\_suka belanja di lazada kualitas barang jelek

**2.3.6. Opinion Word**

Adalah kumpulan kamus kata positif dan negatif dalam Bahasa Indonesia. Dibuat pertama kali oleh Liu dalam penelitiannya kemudian diterjemahkan ke dalam Bahasa Indonesia yang berguna untuk menghitung kata sentimen dalam sebuah kalimat. Dalam kamus kata opinion word terdapat kurang lebih 2400 kata negatif dan 1900 kata positif.

Tabel 2.2 Kumpulan kata positif

A+	ajib	amat	apresiasi	Bakat
acungan	aklamasi	ambisius	asli	Bangga
adaptif	akomodatif	andal	aspirasi	Bantuan
Adil	akurat	anch	asyik	Banyak
afinitas	Alam mimpi	anggun	bagos	Banyak akal
afirmasi	alhamdulillah	Angin sepoi	bagus	Barang baru
Agilely	Allahu akbar	angkat	bahagia	Batu permata
Agung	altruista	antusias	baik	Bebas
Ahli	aman	antusiasme	Baik diposisikan	Masalah
akhirnya	amanah	pik	Baik sekali	Bebas pulsa

Tabel 2.3 Kumpulan kata Negatif

Abnormal	agregator	Amat panas	anarki	Anjlok
Absurd	aib	ambigu	anarkis	Anomali
Acak	ambivalen	anarkisme	antagonis	air terjun
Acak-acakan	akurat	ambivalensi	ancaman	Antagonisme
Acuh	alarm	amoral	anch	Antek
Acuh tak acuh	alasan	amoral	Anch lagi	Anti
Adiktif	Alat permainan	ampun	anehnya	Anti amerika
Adil	alergi	amuk	angkuh	Anti-sirael
Agresi	alegik	Anak nakal	angriness	Anti-kita
Agresif	Amat ketakutan	Anak yatim	anjing	Anti-penduduk

### 2.3.7. Stemming algoritma dari Nazief-Adriani

Stemming: digunakan untuk menghapus imbuhan pada kata. Pada penelitian ini menggunakan algoritma dari Nazief-Adriani dalam menyelesaikan proses stemming. Tahapan pada stemming ini yaitu sebagai berikut:

- a. Pengecekan kata yang belum melewati proses stemming yang terdapat pada kamus. Apabila ada, maka algoritma berhenti .
- b. Menghapus inflection suffixes (“-lah”, “-ku”, “-nya”, dan “-mu”). Namun apabila terdapat (“-tah”, “-lah”, “-pun”, atau “-kah”), maka proses akan diulang yang berguna untuk penghapusan possessive pronouns (“-mu”, “-nya”, “-ku”) pada kamus
- c. Derivation suffixes dihapus (“-kan”, “-i”, “-kan”). Namun algoritma akan berhenti ketika kata masuk dalam kamus dan apabila sebaliknya, maka dilakukan langkah selanjutnya yaitu:
  - Apabila “-an” telah dihapus dan dibagian akhir kata terdapat huruf “k”, maka “-k” dihilangkan. Namun apabila kata tersebut ada pada kamus, maka algoritma berhenti, sedangkan apabila sebaliknya, maka akan dilakukan langkah selanjutnya
  - Akhiran dari kata yang sebelumnya telah melewati tahap penghapusan (“an”, “-kan”, atau “-i”) akan dibawa kembali
- d. Derivational prefix dihapus (“be-”, “me”, “di-”, “te-”, “ke-”, “pe-”, atau “se-”). Apabila pada kamus terdapat kata tersebut, maka algoritma akan berhenti. Apabil sebaliknya, maka dibutuhkan recoding. Apabila selanjutnya terdapat proses seperti yang dijabarkan di bawah ini, maka algoritma berhenti

- Adanya kombinasi pada kata berupa awalan dan akhiran yang tidak diizinkan
  - Terdapat awalan pada kata yang sama persis seperti yang telah dihapuskan sebelumnya
  - Terdapat adanya tiga awalan yang dihapuskan
- e. Apabila langkah-langkah tersebut sudah dilewati, namun kata tersebut tetap tidak termasuk dalam kamus, maka sebelum proses stemming dilakukan kata tersebut akan dikembalikan.

#### **2.3.8. Algoritma klasifikasi**

Klasifikasi adalah proses menemukan model dari sebuah data. Tujuan dari klasifikasi adalah untuk mengambil suatu keputusan dengan memprediksi suatu kasus berdasarkan hasil klasifikasi yang diperoleh. Dalam proses pengklasifikasian terdapat 2 proses yang dilakukan yaitu :

- a. Proses *training* Pada proses ini dilakukan *training set* yang sudah diketahui label-labelnya untuk membangun model.
- b. Proses *testing* Proses ini untuk mengetahui keakuratan model yang dibangun pada proses *training*, umumnya digunakan data yang disebut data *test set* untuk memprediksi label.

#### **2.3.9. Naive Bayes**

*Bayesian* didasarkan pada *teorima bayes* yang memiliki kemampuan hampir serupa dengan *decision tree* dan *neural network*. *Teorema Bayes* adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Bayes merupakan teknik prediksi berbasis *probabilistic* sederhana yang berdasar pada

penerapan *teorema Bayes* (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat atau naif. Dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah “model fitur independen” .

*Bayesian classification* adalah suatu metode pengklasifikasian data dengan model statistik yang dapat digunakan untuk menghitung probabilitas keanggotaan suatu kelas. Metode *Bayesian classification* digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. *Bayesian classification* merupakan salah satu metode yang sederhana yang dapat digunakan untuk data yang tidak konsisten dan data bias. Metode *bayes* juga merupakan metode yang baik dalam mesin pembelajaran berdasarkan data *training* dengan berdasarkan probabilitas bersyarat.

Kaitan antara *Naïve Bayes* dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah hipotesis dalam *teorema Bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi.

Salah satu tugas *data mining* adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. *Naïve Bayes Classifier* merupakan salah satu metoda *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya (Antonius Rachmat C, 2016). Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus Bayes:

$$P(A|B) = (P(B|A) * P(A)) / P(B) \dots\dots\dots (1)$$

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi :

$$P(C_i|D) = (P(D|C_i) * P(C_i)) / P(D) \dots\dots\dots (2)$$

*Naïve Bayes Classifier* atau bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan dari Metode Bayes yang cocok dalam pengklasifikasian teks atau dokumen. Persamaannya adalah:

$$V_{MAP} = \arg \max_{V_j} P(V_j | a_1, a_2, \dots, a_n) \dots\dots\dots (3)$$

Menurut persamaan (3), maka persamaan (1) dapat ditulis:

$$V_{MAP} = \arg \max_{V_j} \frac{P(a_1, a_2, \dots, a_n | P(V_j))}{P(a_1, a_2, \dots, a_n)} \dots\dots\dots (4)$$

Untuk  $P(x_1, x_2, x_3, \dots, x_n)$  nilainya konstan untuk semua kategori ( $V_j$ ) sehingga persamaan dapat ditulis sebagai berikut :

$$V_{MAP} = \arg \max_{V_j} P(X_1, X_2, \dots, X_n | V_j) P(V_j) \dots\dots\dots (5)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \arg \max_{V_j} \prod_{i=1}^n P(X_i | V_j) P(V_j) \dots\dots\dots (6)$$

Keterangan :

$V_j$  = Kategori *Headline*

$J = 1, 2, 3, \dots, n$ . Dimana dalam penelitian ini

$j_1$  = kategori *headline* sentimen positif

$j_2$  = kategori *headline* sentimen negatif

$j_3$  = kategori *headline* sentiment netral

$P(x_i|V_j)$  = Probabilitas  $x_i$  pada kategori  $V_j$

$V_j P(V_j)$  = Probabilitas dari  $V_j$



Untuk  $P(V_j)$  dan  $P(x_i|V_j)$  dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \dots \dots \dots (7)$$

Jika  $P(V_j)$  sudah ditentukan maka hitung jumlah dokumen setiap kategori  $j$  dan jumlah dokumen dari semua kategori dengan menggunakan persamaan 7. diatas.

$$P(x_i|V_j) = \frac{nk+1}{nk+[konstanta]} \dots \dots \dots (8)$$

Keterangan :

$|docs\ j|$  = jumlah dokumen setiap kategori  $j$

$|contoh|$  = jumlah dokumen dari semua kategori

$nk$  = jumlah frekuensi kemunculan setiap kata

$n$  = jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$  = jumlah semua kata dari semua kategori

Jika  $P(x_i|V_j)$  sudah ditentukan maka hitung jumlah frekuensi kemunculan setiap kata ditambah 1 dan jumlah frekuensi kemunculan kata dari setiap kategori dengan menggunakan persamaan 8. diatas.

### 2.3.10. K-Nearest Neighbor (K-NN)

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Widiarsana, O et al., 2011). Algoritma K-Nearest Neighbor (KNN) adalah metode yang digunakan untuk mengelompokkan objek berdasarkan contoh pelatihan terdekat di ruang fitur. K-Nearest Neighbor merupakan jenis yang paling dasar dari contoh *based learning* atau *lazy learning*

juga termasuk kelompok *instance-based learning*. K-Nearest Neighbor dilakukan dengan mencari kelompok objek dalam data training yang paling dekat (mirip dengan objek pada data baru atau data *testing*). (Krtati,2014).

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (*data training*), diantaranya *Euclidean distance* dan *manhattan distance* (*city block distance*) yang sering digunakan adalah *euclidean distance*, yaitu :

$$d(x, y) = \sqrt{\sum_i^n (a_i - b_i)^2 + (a_2 - b_2)^2, \dots, \dots, (a_n - b_n)^2} \dots \dots \dots (9)$$

Dimana  $a = a_1, a_2, \dots, \dots, a_n$  dan  $b = b_1, b_2, \dots, \dots, b_n$  mewakili  $n$  nilai *atribut* dari dua *record*. Adapun langkah-langkah untuk menghitung algoritma K-Nearest Neighbor antara lain :

1. Menentukan Parameter K (jumlah tetangga terdekat).
2. Menghitung kuadrat jarak *eucliden* (*queri instance*) masing –masing objek terhadap data sampel yang diberikan.
3. Kemudian mengurutkan jarak tersebut ke dalam kelompok yang mempunyai *Euclid* terkecil (mengurutkan hasil no 2 secara *ascending*).
4. Mengumpulkan kategori Y (Klasifikasi K-NN) berdasarkan nilai K atau ambil data tetangga terdekat.

Dengan menggunakan kategori K-NN yang paling mayoritas maka akan menghasilkan kelas data baru (prediksi). Nilai k yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai k yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *Simple Unweighted Voting*.

### 2.3.11. Particle Swarm Optimization (PSO)

*Particle Swarm Optimization* (PSO) diperkenalkan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995, merupakan algoritma optimasi yang meniru proses yang terjadi dalam kehidupan populasi burung (*flock of bird*) dan ikan (*school of fish*) dalam bertahan hidup. Sejak diperkenalkan pertama kali, algoritma PSO berkembang cukup pesat, baik dari sisi aplikasi maupun dari sisi pengembangan metode yang digunakan pada algoritma tersebut (Haupt, R.L. & Haupt, S.E. 2004). Oleh sebab hal tersebut, mereka mengategorikan algoritma sebagai bagian dari kehidupan *rekayasa/buatan Artificial Life*. Algoritma ini juga terhubung dengan komputasi *evolusioner*, algoritma genetik dan pemrograman evolusioneri (Jatmiko et al. 2010).

Adapun parameter yang digunakan dalam proses algoritma PSO adalah sebagai berikut:

1. Jumlah partikel (*Number of particles*) merupakan faktor yang dianggap sangat penting dalam melakukan penyelesaian masalah.
2. Bobot Inersia (*Inertia Weight*), memainkan peran yang sangat penting dalam kecepatan partikel dari algoritma PSO.
3. Faktor pembelajaran (*Learning factors*). Parameter  $c_1$  merupakan pengakuan koefisien, sedangkan  $c_2$  adalah komponen social. Hal ini tidak terlalu penting untuk perilaku konvergensi dari algoritma PSO
4. Rentang dan dimensi partikel (*Range and dimension of particles*). Dimensi partikel dan rentang ditentukan berdasarkan masalah yang dioptimalkan.

5. Kecepatan (*Velocity*) merupakan perubahan maksimum pada setiap partikel, dapat diambil selama iterasi yang didefinisikan sebagai kecepatan maksimum.
6. Menghentikan kondisi (*Stopping condition*) merupakan salah satu cara apabila kriteria yang dicari sudah tercapai.

Istilah yang digunakan dalam penerapan algoritma PSO sebagai berikut:

1. *Swarm*: populasi dari sekawanan partikel
2. *Particle*: individu pada suatu *swarm*. Setiap partikel mempresentasikan suatu solusi dari permasalahan yang diselesaikan.
3. *Pbest*: suatu partikel yang menunjukkan posisi terbaik.
4. *Gbest*: posisi terbaik dari seluruh partikel yang ada dalam suatu *swarm*.
5. *Velocity*: kecepatan yang dimiliki oleh setiap partikel dalam menentukan arah perpindahan suatu partikel untuk memperbaiki posisi semula.
6.  $c_1$  dan  $c_2$ :  $c_1$  merupakan konstanta pembelajaran kognitif, dan  $c_2$  konstanta pembelajaran sosial.

Untuk memulai algoritma PSO, kecepatan awal (*velocity*) dan posisi awal (*position*) ditentukan secara *random*. Kemudian proses pengembangannya sebagai berikut:

- a. Asumsikan bahwa ukuran kelompok atau kawanan (jumlah partikel) adalah  $N$ . Kecepatan dan posisi awal pada tiap partikel dalam  $N$  dimensi ditentukan secara *random* (acak).
- b. Hitung kecepatan dari semua partikel. Semua partikel bergerak menuju titik optimal dengan suatu kecepatan. Awalnya semua kecepatan dari partikel diasumsikan sama dengan nol, set iterasi  $i = 1$ .
- c. Nilai *fitness* setiap partikel ditaksir menurut fungsi sasaran (*objective function*) yang ditetapkan. Jika nilai *fitness* setiap partikel pada lokasi saat ini lebih baik dari *Pbest*, maka *Pbest* diatur untuk posisi saat ini.

- d. Nilai *fitness* partikel dibandingkan dengan *Gbest*. Jika *Gbest* yang terbaik maka *Gbest* yang di-*update*.
- e. Persamaan (10) dan (11) ditunjukkan di bawah ini untuk memperbaharui (*update*) kecepatan (*velocity*) dan posisi (*position*) setiap partikel.

$$V_{id}^{k+1} = w \times V_{id}^k + c1 \times rand1 \times (P_{id} - X_{id}) + c2 \times rand2 \times (G_{id} - X_{id}) \dots (10)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \dots (11)$$

Dimana:

$V_{id}$  = komponen kecepatan individu ke  $i$  pada  $d$  dimensi

$X_{id}$  = posisi individu  $i$  pada  $d$  dimensi

$\omega$  = parameter *inertia weight*

$c1$   $c2$  = konstanta akselerasi (*learning rate*), nilainya antara 0 sampai 1  $rand1$ ,

$rand2$  = parameter *random* antara 0 sampai 1

$P_{id}$  = *Pbest (local best)* individu  $i$  pada  $d$  dimensi

$G_{id}$  = *Gbest (global best)* pada  $d$  dimensi

- f. Cek apakah solusi yang sekarang sudah *konvergen*. Jika posisi semua partikel menuju ke satu nilai yang sama, maka ini disebut *konvergen*. Jika belum *konvergen* maka langkah 2 diulang dengan memperbarui iterasi  $i = i + 1$ , dengan cara menghitung nilai baru dari  $P_{best,j}$  dan  $G_{best}$ . Proses iterasi ini dilanjutkan sampai semua partikel menuju ke satu titik solusi yang sama. Biasanya akan ditentukan dengan kriteria penghentian (*stopping criteria*), misalnya jumlah selisih solusi sekarang dengan solusi sebelumnya sudah sangat kecil.

g. Ada 2 aspek penting dalam memilih kondisi berhenti yaitu:

1. Kondisi berhenti tidak menyebabkan PSO *convergent premature* (memusat sebelum waktunya) dimana solusi tidak optimal yang didapat.
2. Kondisi berhenti harus melindungi dari kondisi *oversampling* pada nilainya, jika kondisi berhenti memerlukan perhitungan yang terus menerus maka kerumitan dari proses pencarian akan meningkat.

Beberapa kondisi berhenti yang dapat dipakai dalam *Particle Swarm Optimization* adalah berhenti ketika jumlah iterasi telah mencapai jumlah iterasi maksimum yang diperbolehkan, berhenti ketika solusi yang diterima ditemukan, Berhenti ketika tidak ada perkembangan setelah beberapa iterasi.

#### **2.3.12. Cross Validation**

*Cross Validation* merupakan teknik validasi dengan cara membuat data terbagi menjadi dua bagian secara acak, dimana data tersebut dibagi menjadi data training dan data testing. Data training adalah data yang akan dipakai dalam melakukan pembelajaran. Sedangkan data testing ialah data yang digunakan sebagai pengujian dari model yang telah dibuat. Dengan menggunakan split validation maka akan dilakukan percobaan pada data training berdasarkan pada split ratio yang telah ditentukan. Setelah itu sisa dari split ratio pada data training digunakan untuk data testing, misalkan jika split ratio pada data training 80% maka data testing sebesar 20% begitupun seterusnya (I. H. Witten, E. Frank, and M. A. Hall, 2011 ). Untuk lebih jelasnya dapat melihat Tabel dibawah :

Tabel 2.4. Ilustrasi Split Validation

Training 90%										Test 10%
Training 80%										Test 20%
Training 70%										Test 30%
Training 60%										Test 40%
Training 50%										Test 50%
Training 40%										Test 60%
Training 30%										Test 70%
Training 20%										Test 80%
Training 10%										Test 90%

## BAB III

### METODE PENELITIAN

#### 3.1. Jenis, Sifat, dan Pendekatan Penelitian

Twitter sebagai media sosial dimana penggunanya bebas menulis *tweet* yang mereka inginkan. Terdapat banyak metode untuk analisis sentiment pada Twitter, pada penelitian ini metode yang digunakan adalah k-NN, Naive Bayes dan PSO. Penelitian dimulai dengan *crawling* data calon presiden Indonesia tahun 2019 yaitu Jokowi dan Prabowo menggunakan Rapidminer. Kemudian data dilakukan *cleaning* data, dan labeling secara manual. Kemudian metode k-NN dan Naive Bayes digunakan untuk mengklasifikasi sentiment masyarakat, setelah itu kedua metode tersebut akan di optimalisasi dengan menggunakan metode PSO. Hasil akurasi dari metode tersebut akan dibandingkan tingkat akurasinya kemudian dapat dipilih metode mana yang terbaik.

#### 3.2. Metode Pengumpulan Data

Dalam penelitian sentiment analisis terhadap calon presiden tentunya dibutuhkan data sebagai komponen utama sebuah penelitian. Pengumpulan data penelitian dilakukan dengan menggunakan *software* Rapidminer. RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan dengan nama yang sama yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analitik prediktif. Dengan menjalankan proses di bawah ini kita dapat mengambil data *tweet* :





Gambar 3.1 Proses pengumpulan data Twitter.

Dengan operator Pencarian Twitter, kita dapat mencari *tweet* untuk proses penelitian. Sebelum melakukan proses pengambilan data, kita harus mengatur parameter pada Search Twitter, seperti yang terlihat pada gambar berikut:



Gambar 3.2 Parameter Search Twitter

Parameter connection digunakan untuk mengkoneksikan Rapidminer dengan Twitter, untuk dapat terkoneksi dengan Twitter kita harus mempunyai akun Twitter terlebih dahulu. Untuk mempunyai akun Twitter anda bisa registrasi di twitter.com. Parameter query digunakan untuk melakukan pencarian tweet dalam database Twitter, contoh query yang akan dicari adalah jokowi. Parameter limit adalah batasan jumlah *tweet* yang akan dicari. Apabila kita mengisi dengan angka 10 maka hasil *tweet* yang akan diperoleh berjumlah 10 record.

Pada Gambar 3.2 terdapat proses Remove Duplicate yang digunakan untuk menghapus duplikasi *tweet* yang ada, sehingga data yang didapat lebih bervariasi.

Pada gambar 3.2 terdapat proses Write Excel yang digunakan untuk menuliskan *tweet* yang didapat kedalam file excel (.xlsx). Dari proses, tersebut dapat dilihat hasil pencarian *tweet* yang didapat pada gambar sebagai berikut:

	B	C	D	E	F	G	H	I	J	K	L	M
1	From User	From User To User	To User	Language	Source	Text	Geo-Loac	Geo-Location	Longitude	Retweet	Count	Id
2	Joko Widodo	_36987170	-1	id	ca	ca tweet "Tidak mungkin bangsa Indonesia menjadi bang				4257	0	1118041064670230561
3	Joko Widodo	_36987170	-1	id	ca	ca tweet "Aku pernah panggikanMu, ya Allah, aku pernah				6060	0	1118150200172826025
4	Joko Widodo	_36987170	-1	id	ca	ca tweet "Saya ajak makan bakso, mau. Saya ajak buah				8695	0	1118098230574535232
5	Atbi Kartawana	_59029935	-1	id	ca	ca tweet "RT @dika5710: Bang Indonesia itu baw				1214	0	1118101050581730881
6	akulifre	_38148033	-1	id	ca	ca tweet "RT @Kakabum: Dik dike ingatkan oleh bap				1044	0	1118101048810395408
7		_51272913	0	id	ca	ca tweet "RT @Kanghwa: Eke bawakanlah, Dendur				2800	0	1118101044121228073
8	Sukotirno	_126317056	4464	id	ca	ca tweet "RT @Kakabum: Hari hari yang memanggikan				1075	0	1118101037292826024
9	@suzanb402	_736830200	-1	id	ca	ca tweet "RT @Kakabum: Dik dike ingatkan oleh bap				1044	0	11181010312261821044
10	Arastayof	_11310027083	20	id	ca	ca tweet "Hari ini waktu kita pilih jenderalArastayof				0	0	111810102256487728
11	Israng p.k	_381761078	-1	id	ca	ca tweet "RT @alubali: Prabowo sudah tidak akan men				222	0	111810102225957107

Gambar 3.3 Hasil Pencarian data Twitter

Pada gambar 3.3 terlihat hasil pencarian Twitter yang didapat dengan menggunakan Rapidminer berupa From-User, Language, Source, Text, dll. Dari semua kolom data yang didapatkan hanya kolom text yang digunakan untuk penelitian. Text dari hasil pengambilan data dapat dilihat pada gambar berikut :

Tidak mungkin bangsa Indonesia menjadi bangsa yang maju, menjadi negara yang kompetitif, menjadi Poros Maritim Dunia, bangsa yang berdaulat di bidang pangan, kalau Infrastruktur - jalan, jembatan, pelabuhan, bandara, bendungan dan irigasi, dan sebagainya - tidak memadai. <a href="https://t.co/wUKINCeQ3">https://t.co/wUKINCeQ3</a>
Aku pernah panggikanMu, ya Allah, aku pernah panggikanMu.
Tidak ada sekutu bagiMu, aku pernah panggikanMu.
Sesungguhnya segala puji, nikmat dan kerajaan bagiMu. Tidak ada sekutu bagiMu. <a href="https://t.co/K6C3r8Xlv">https://t.co/K6C3r8Xlv</a>
Saya ajak makan bakso, mau. Saya kirim buah duren kesukaannya, mau.
Dari teman dekat adk saya, kemudian jadilah ia teman hidup saya. <a href="https://t.co/008Gx#13mT">https://t.co/008Gx#13mT</a>

Gambar 3.4 *Tweet* hasil pencarian data

### 3.3. Preprocessing Data

Data *tweet* yang telah diambil dari Twitter masih berupa data mentah maka dari itu dilakukan tahap preprocessing untuk mendapatkan data bersih agar dapat

diproses ke tahap selanjutnya. Tahapan yang dilakukan adalah transform case, cleaning data, normalisasi dan stemming.

```
# ===== DATA CLEANSING
# Work with Corpus
tweet.corpus = VCorpus(VectorSource(Data_joko_Widodo$Text))
inspect(tweet.corpus[[1]])
# Import external function
source(file = 'D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Function Helper/Cleaning.R')
# TRANSFORM TO LOWER CASE
tweet.corpus = tm_map(tweet.corpus, content_transformer(tolower))
# FILTERING - CUSTOM CLEANSING FUNCTIONS
tweet.corpus = tm_map(tweet.corpus, content_transformer(removeURL))
tweet.corpus = tm_map(tweet.corpus, content_transformer(unescapeHTML))
tweet.corpus = tm_map(tweet.corpus, content_transformer(removeMention))
tweet.corpus = tm_map(tweet.corpus, content_transformer(removeCarriage))
tweet.corpus = tm_map(tweet.corpus, content_transformer(removeEmoticon))
tweet.corpus = tm_map(tweet.corpus, content_transformer(removeInvoice))
# Remove additional symbols to white space
tweet.corpus = tm_map(tweet.corpus, toSpace, "[[:punct:]]") # punctuation
tweet.corpus = tm_map(tweet.corpus, toSpace, "[[:digit:]]") # numbers
# Eliminate extra white spaces
tweet.corpus = tm_map(tweet.corpus, stripWhitespace)
# Check the final result
inspect(tweet.corpus[[1]])
# SPELLING NORMALIZATION
spell.lex = read.csv(file = 'D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Data Helper/colloquial-Indonesian-lexicon.txt',
                    header = TRUE,
                    sep = ',',
                    stringsAsFactors = FALSE)
# Import external function
source(file = 'D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Function Helper/Cleaning.R')
tweet.corpus = tm_map(tweet.corpus, spell.correction, spell.lex)
# STEMMING WORDS
# Import external function
source(file = 'D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Function Helper/Cleaning.R')
tweet.corpus = tm_map(tweet.corpus, content_transformer(stemming))
```

Gambar 3.5 Preprocessing

### 3.3.1. Transform case

Tabel 3.1 Transform case

Fungsi	Kegunaan
# TRANSFORM TO LOWER CASE tweet.corpus = tm_map(tweet.corpus, content_transformer(tolower))	Mengubah semua huruf menjadi kecil

Tahapan ini merupakan tahapan untuk menyamaratakan bentuk huruf dalam kalimat, dari kalimat yang memiliki huruf kapital diubah menjadi huruf kecil dan disamakan mulai dari a sampai z.

tidak mungkin bangsa Indonesia menjadi bangsa yang maju, menjadi negara yang kompetitif, menjadi poros maritim dunia, bangsa yang berdaulat di bidang pangan, kalau infrastruktur - jalan, jembatan, pelabuhan, bandara, bendungan dan irigasi, dan sebagainya - tidak memadai. <a href="https://t.co/wukinogq3">https://t.co/wukinogq3</a>
aku penuhi panggilanmu, ya Allah, aku penuhi panggilanmu. tidak ada seketu bagimu, aku penuhi panggilanmu. sesungguhnya segala puji, nikmat dan kerajaan bagimu. tidak ada seketu bagimu. <a href="https://t.co/kbc3tr8vly">https://t.co/kbc3tr8vly</a>
saya ajak makan bakso, mau. saya kirim buah durian kesukaannya, mau. dari teman dekat adik saya, kemudian jadilah ia teman hidup saya. <a href="https://t.co/008kpr13mt">https://t.co/008kpr13mt</a>
rt @bakarsmith: astaghfirullah ada dua dusta di sini: prabowo naik haji tahun 1991, dan sudah masuk kabah. 1.(bulan haji pintu ka'bah t...

Gambar 3.6 Hasil transform case

Dapat dilihat bahwa semua huruf yang berada pada kalimat berubah menjadi kecil. Tahapan ini dapat mempermudah dan mempercepat proses klasifikasi agar mendapatkan waktu terbaik.

### 3.3.2. Cleaning data

Tabel 3.2 Cleaning data

Fungsi	Kegunaan
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(removeURL))</code>	Menghilangkan URL
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(unescapeHTML))</code>	Menghapus ASCII dan kode hexadesimal pada HTML
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(removeMention))</code>	Menghapus mention
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(removeCarriage))</code>	Menghapus carriage
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(removeEmoticon))</code>	Menghapus emoticon
<code>tweet.corpus = tm_map(tweet.corpus, content_transformer(removeInvoice))</code>	Menghapus invoice
<code>tweet.corpus = tm_map(tweet.corpus, toSpace, "[[:punct:]]")</code>	Menghilangkan punctuation
<code>tweet.corpus = tm_map(tweet.corpus, toSpace, "[[:digit:]]")</code>	Menghapus nomor
<code>tweet.corpus = tm_map(tweet.corpus, stripWhitespace)</code>	Menghapus spasi berlebih

Pada tabel diatas terdapat fungsi-fungsi yang digunakan untuk menghilangkan simbol dan angka yang tidak diperlukan. Dari hasil proses di atas dapat dilihat pada gambar dibawah ini:

tidak mungkin bangsa Indonesia menjadi bangsa yang maju menjadi negara yang kompetitif menjadi poros maritim dunia bangsa yang berdaulat di bidang pangan kalau infrastruktur jalan jembatan pelabuhan bandara bendungan dan irigasi dan sebagainya tidak memadai
aku penuhi panggilanmu ya allah aku penuhi panggilanmu tidak ada sekutu bagimu aku penuhi panggilanmu sesungguhnya segala puji nikmat dan kerajaan bagimu tidak ada sekutu bagimu
sayaajak makan bakso mau saya kirim buah durian kesukaanmu mau dari teman dekat adik saya kemudian jadilah ia teman hidup saya
astaghfirullah ada dua dusta di sini Prabowo naik haji tahun dan sudah masuk kabah bulan haji pintu kabah t
dikit dikit lapor dikit dikit lapor kok dikit dikit demokrasi terancam apabila tak siap beda pendapat yang rubah foto

Gambar 3.7 Hasil cleaning data.

Setelah fungsi diatas dijalankan maka akan secara otomatis menghilangkan simbol dan angka pada kalimat. Tahapan ini sangat berpengaruh dan dilakukan agar mendapatkan data yang valid untuk diolah pada tahap selanjutnya.

### 3.3.3. Normalisasi

Tabel 3.3 Normalisasi

Fungsi	Kegunaan
<pre>spell_lex = read.csv(file = "D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Data Helper/colloquial- Indonesian-lexicon.txt", header = TRUE, sep = ";", stringsAsFactors = FALSE) # Import external function sourcefile = "D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Function Helper/Cleansing_R") tweet.corpus = tm_map(tweet.corpus, spell.correction, spell_lex)</pre>	Melakukan perbaikan pada kata-kata yang mengalami kesalahan penulisan

Normalisasi digunakan untuk melakukan perbaikan pada kata-kata yang mengalami kesalahan penulisan atau typo, seperti kata "bget" diubah menjadi "banget", dan lain sebagainya.

1	slang, formal
2	briziek, berisiek
3	woww, wow
4	aminn, amin
5	met, selamat
6	netaaa, menetas
7	keberpa, keberapa
8	eeehhhh, eh
9	kata2nyaaa, kata-katanya
10	hello, halo

Gambar 3.8 Kamus normalisasi

Ada lebih dari 5.000 kamus perbaikan kata yang digunakan untuk normalisasi. Proses normalisasi adalah proses yang sangat penting untuk peningkatan akurasi. Seperti yang kita ketahui masyarakat Indonesia cenderung menuliskan kata-kata pada sosial media dengan bahasa gaul ataupun singkatan, apabila kata tersebut tidak dibuat menjadi kata baku maka hasil akurasi akan menjadi sedikit. Berikut contoh hasil normalisasi:

tidak mungkin bangsa Indonesia menjadi bangsa yang maju menjadi negara yang kompetitif menjadi poros maritim dunia bangsa yang berdaulat di bidang pangan kalau infrastruktur jalan jembatan pelabuhan bandara bendungan dan irigasi dan sebagainya tidak memadai
aku penuh pengalamannya ya Allah aku penuh pengalamannya tidak ada sekutu bagimu aku penuh pengalamannya sesungguhnya segala puji nikmat dan kerajaannya tidak ada sekutu bagimu
saya ajak makan bakso mau saya kirim buah durian kesukaannya mau dari teman dekat adik saya kemudian jadilah ia teman hidup saya
astaghfirullah ada dua dusta di sini prabowo naik haji tahun dan sudah masuk kabah bulan haji pinta kabah 1 dikit dikit lapor dikit dikit lapor lapor kok dikit dikit demokrasi terancam apabila tak siap beda pendapat yang rubah foto

Gambar 3.9 Hasil Normalisasi

Pada contoh di atas kebetulan tidak ada kata yang perlu diperbaiki sehingga tidak ada perubahan dalam kalimatnya.

### 3.3.4. Stemming

Tabel 3.4 Stemming

Fungsi	Kegunaan
source(file = "D:/kuliah/KULIAH S2/Tests/848 345/program R/Helpers/Function Helper/Cleaning.R") tweet.corpus = tm_map(tweet.corpus, content_transformer(stemming))	Menghapus imbuhan pada kata

Stemming digunakan untuk menghapus imbuhan pada kata. Pada penelitian ini menggunakan algoritma dari Nazief-Adriani dalam menyelesaikan proses stemming. Dari fungsi stemming dihasilkan data sebagai berikut:

tidak mungkin bangsa Indonesia menjadi bangsa yang maju menjadi negara yang kompetitif menjadi poros maritim dunia bangsa yang diulat di bidang pangan kalau infrastruktur jalan jembatan pelebuhan bandara bendungan dan irigasi dan sebagainya tidak memadai
aku penuhi panggilanmu ya allah aku penuhi panggilanmu tidak ada sekutu bagimu aku penuhi panggilanmu sungguh segala puji nikmat dan raja bagimu tidak ada sekutu bagimu
saya ajak makan bakso mau saya kirim buah durian suka mau dari teman dekat edik saya kemudian jadi ia teman hidup saya
astagfirullah ada dua dusta di sini grabowo nak haji tahun dan sudah masuk kabah bulan haji pintu kabah t dikit dikit lapor dikit dikit lapor dikit dikit demokrasi terancam apabila tak siap beda pendapat yang ubah foto

Gambar 3.10 Hasil Stemming

Setelah proses stemming dapat dilihat beberapa kata berimbuhan berubah menjadi kata dasar seperti “berbeda” menjadi “beda”.

### 3.3.5. Proses penfilan sentiment

Setelah melalui proses preprocessing dilanjutkan dengan proses penilaian dengan cara memberikan nilai positif dan negatif pada kalimat dengan cara statistikal learning. Untuk memberikan nilai dibutuhkan dataset kata latih positif dan negatif untuk pembelajaran komputer. Seperti yang terlihat pada gambar dibawah ini:

1	+	1	abnormal
2	acungan jempol	2	absurd
3	adaptif	3	acak
4	adil	4	acak-acakan
5	efinitas	5	acuh
6	afirmasi	6	acuh tak acuh
7	agilely	7	adiktif
8	agung	8	adil
9	ahli	9	agresi
10	ahlinya	10	agresif

Gambar 3.11 Contoh kata positif dan negatif

Jumlah daftar kata positif sebanyak 1182 kata dan kata negatif berjumlah 2402 kata. Kalimat akan otomatis mendapatkan penilaian kelipatan 1 apabila terdapat kata positif dalam kalimat, begitu pula apabila terdapat kata negatif maka akan otomatis mendapatkan penilaian kelipatan -1. Dalam tahap ini penilaian

dilakukan dengan menggunakan bahasa R, dengan menjalankan kode program sebagai berikut:

```
# ===== LEXICON-BASED SENTIMENT SCORING
pos = readLines("D:/Kuliah/KULIAH S2/Tesis/BAB 345/program
R/Helpers/Data Helper/s-pos.txt")
neg = readLines("D:/Kuliah/KULIAH S2/Tesis/BAB 345/program
R/Helpers/Data Helper/s-neg.txt")
df.tweet = data.frame(text = sapply(tweet.corpus, as.character),
stringsAsFactors = FALSE)
# Negation Handling
negasi = scan("D:/Kuliah/KULIAH S2/Tesis/BAB 345/program R/Helpers/Data
Helper/negatingword.txt",
what = 'character')
senti = read.csv("D:/Kuliah/KULIAH S2/Tesis/BAB 345/program
R/Helpers/Data Helper/sentiwords_id.txt",
sep = ";",
header = FALSE) %>%
mutate(words = as.character(V1),
score = as.numeric(V2)) %>%
select(c('words','score'))
booster = read.csv("D:/Kuliah/KULIAH S2/Tesis/BAB 345/program
R/Helpers/Data Helper/boosterwords_id.txt",
sep = ";",
header = FALSE) %>%
mutate(words = as.character(V1),
score = as.numeric(V2)) %>%
select(c('words','score'))
# Import external function
source(file = "D:/Kuliah/KULIAH S2/Tesis/BAB 345/program
R/Helpers/Function Helper/Lexicon-Based Scoring Analysis.R")
results = scores.sentiment(df.tweet$Text, senti, negasi)
```



```

set.seed(2211)
random.tweet = sample(x = df.tweet$Text,
                      size = 10)
head(scores.sentiment(random.tweet, senti, negasi),5)
# Convert score to sentiment classes
results$class = as.factor(ifelse(results$score < 0, 'Negative',
                                ifelse(results$score == 0, 'Neutraf','Positive')))

```

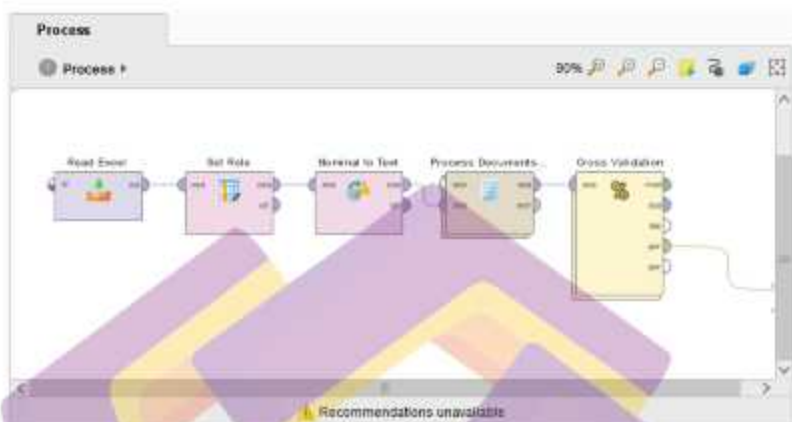
Gambar 3.12 Sentiment scoring

Sebelum dilakukan proses penilaian, kalimat akan diubah menjadi huruf kecil dan penghilangan simbol yang dianggap noise. Setelah kalimat tersebut bersih dilanjutkan dengan proses penilaian, dimana kalimat positif akan bernilai 0 atau lebih, sedangkan kalimat negatif akan bernilai -1 atau lebih. Contoh penilaian dapat dilihat pada tabel dibawah ini :

Tabel 3.5 sentiment scoring

Text	Sentiment
tidak mungkin bangsa indonesia menjadi bangsa yang maju menjadi negara yang kompetitif menjadi poros maritim dunia bangsa yang daulat di bidang pangan kalau infrastruktur jalan jembatan pelabuhan bandara bendungan dan irigasi dan sebagainya tidak memadai	Negative
aku penuhi panggilanmu ya allah aku penuhi panggilanmu tidak ada sekutu bagimu aku penuhi panggilanmu sungguh segala puji nikmat dan raja bagimu tidak ada sekutu bagimu	Neutral
saya ajak makan bakso mau saya kirim buah durian suka mau dari teman dekat adik saya kemudian jadi ia teman hidup saya	Positive
astaghfirullah ada dua dusta di sini prabowo naik haji tahun dan sudah masuk kabah bulan haji pintu kabah t	Negative
dikit dikit lapor dikit dikit lapor lapor kok dikit dikit demokrasi terancam apabila tak siap beda pendapat yang ubah foto	Negative

### 3.4. Klasifikasi sentiment analysis

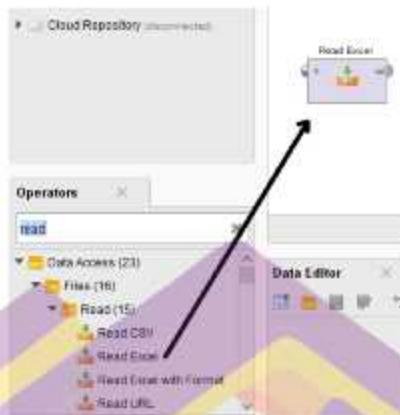


Gambar 3.13 Klasifikasi sentiment analysis

Proses klasifikasi dimulai dari retrieve historical sentiment yaitu proses mengakses data yang telah di beri sentiment positif dan negatif dan memuatnya kedalam proses. Selanjutnya dilanjutkan dengan set role, operator ini digunakan untuk mengubah peran atribut dalam hal ini atribut sentiment digunakan sebagai label. Kemudian dilanjutkan nominal to text, operator ini mengubah jenis atribut nominal yang dipilih untuk teks. Ini juga memetakan semua nilai atribut ini ke nilai string yang sesuai. Selanjutnya preprocessing data, dan diakhiri dengan cross validation.

#### 3.4.1. Read Excel

Digunakan untuk memanggil data file yang memiliki ekstensi xlsx, tool ini dapat mempermudah memanggil data. Adapun langkah-langkah menjalankan fitur pada Rapidminer sebagai berikut.



Gambar 3.14 Tahapan read excel

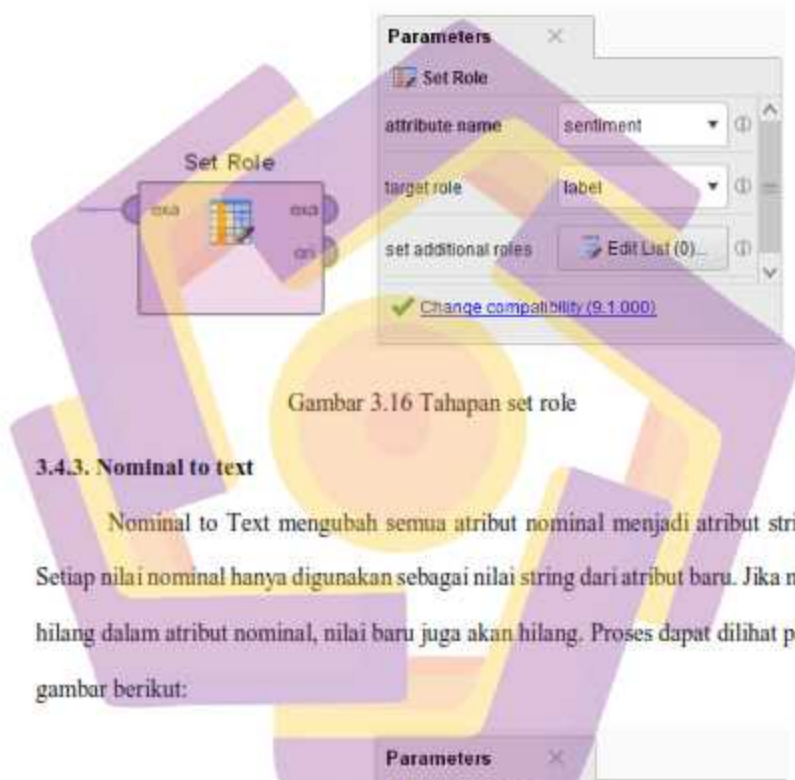
Pada fitur *read Excel* drag ke bagian lembar kerja yang kosong untuk menampilkan *tool read Excel* ke layar penelitian. Tahapan berikutnya yaitu *Excel file* dari fitur *read Excel* dengan klik gambar folder, adapun langkahnya adalah sebagai berikut:



Gambar 3.15 Open file

### 3.4.2. Set role

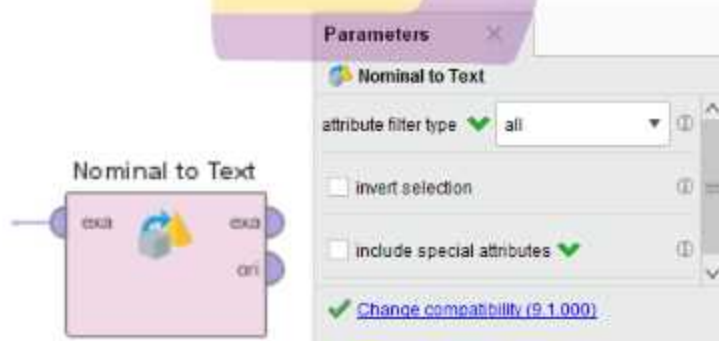
Set role merupakan proses penentuan atribut dalam data, atribut yang digunakan adalah atribut sentiment dan gunakan sebagai label, seperti pada gambar di bawah ini:



Gambar 3.16 Tahapan set role

### 3.4.3. Nominal to text

Nominal to Text mengubah semua atribut nominal menjadi atribut string. Setiap nilai nominal hanya digunakan sebagai nilai string dari atribut baru. Jika nilai hilang dalam atribut nominal, nilai baru juga akan hilang. Proses dapat dilihat pada gambar berikut:



Gambar 3.17 Tahapan Nominal to text

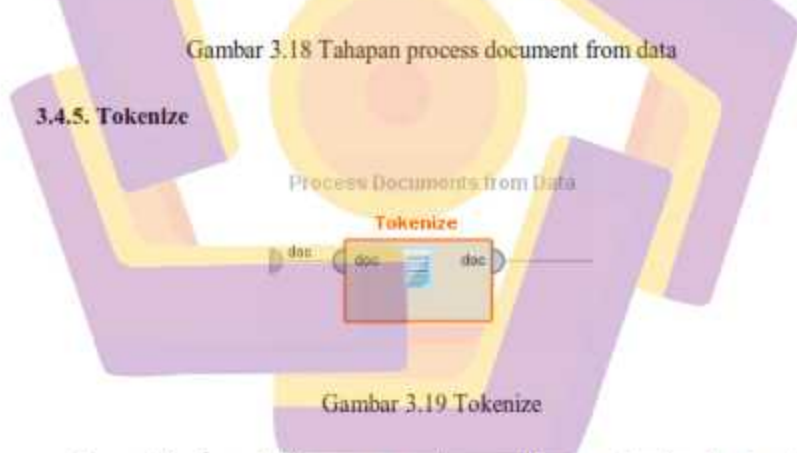
### 3.4.4. Process document from data

Tahapan Process document from data adalah tahapan proses-proses seperti tokenize, filter tokens by length, dan filter stopwords indonesia berada.



Gambar 3.18 Tahapan process document from data

### 3.4.5. Tokenize



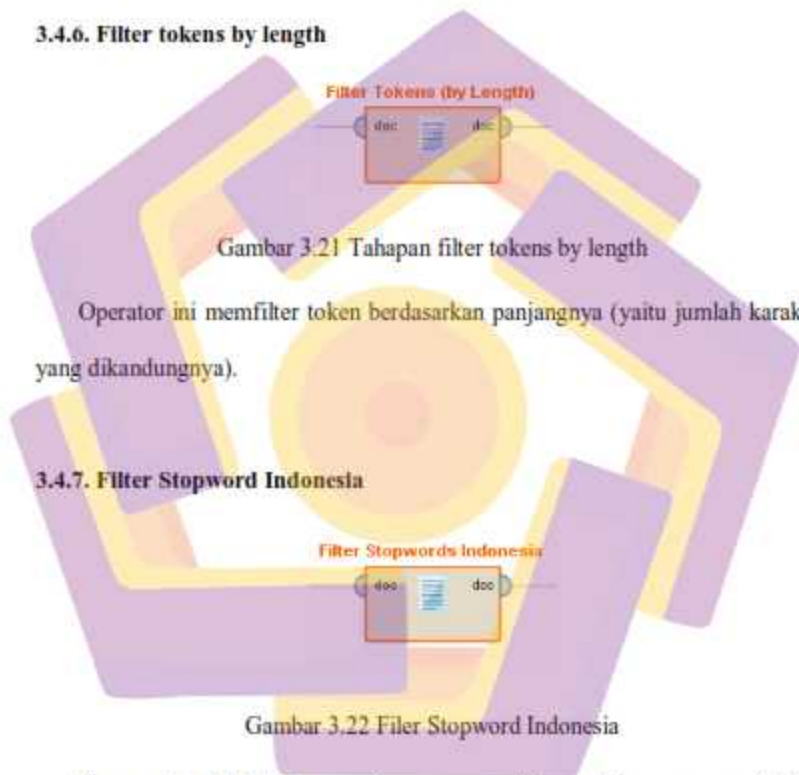
Gambar 3.19 Tokenize

Proses tokenize adalah proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses memisahkan setiap kata yang menyusun suatu dokumen. Pada umumnya setiap kata teridentifikasi atau terpisahkan dengan kata yang lain oleh karakter spasi, sehingga proses tokenizing mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata.

tidak	mungkin	bangsa	Indonesia	menjadi	bangsa	yang	maju
menjadi	negara	yang	kompetitif	menjadi	poros	martim	dunia
bangsa	yang	daulat	di	bidang	pangan	kalah	infrastruktur
jalan	jembatan	pelabuhan	bandara	bendungan	dan	irigasi	dan
sebagainya	tidak	memadai					

Gambar 3.20 Contoh hasil tokenizing

### 3.4.6. Filter tokens by length



Gambar 3.21 Tahapan filter tokens by length

Operator ini memfilter token berdasarkan panjangnya (yaitu jumlah karakter yang dikandungnya).

### 3.4.7. Filter Stopword Indonesia

Filter Stopwords Indonesia



Gambar 3.22 Filer Stopword Indonesia

Stop words adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Filter Stopword digunakan untuk menghapus kata yang tidak memiliki makna pada kalimat. Seperti yang terlihat pada gambar dibawah:

1	ada
2	adalah
3	adanya
4	adapun
5	agak
6	agakny
7	agar
8	akan
9	akankah
10	akhir

Gambar 3.23 Kamus Stopword Indonesia

Kamus stopwords Indonesia berjumlah 759 kata, dengan menggunakan kamus ini dihasilkan data sebagai berikut:

bangsa	Indonesia	bangsa	maju	negara	kompetitif	poros
maritim	Dunia	bangsa	daulat	bidang	pangan	infrastruktur
jalan	jembatan	pelabuhan	bandara	bendungan	irigasi	memadai

Gambar 3.24 Contoh hasil Filter stopwords

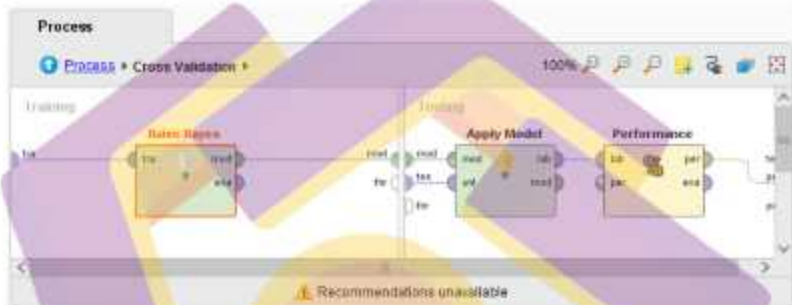
### 3.4.8. Cross Validation



Gambar 3.25 Tahapan cross validation

Cross validation digunakan untuk memperkirakan seberapa akurat suatu model (dipelajari oleh Operator pembelajaran tertentu) dan akan ditampilkan dalam

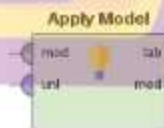
praktik Cross validation adalah Operator yang bersarang, memiliki dua subproses: subproses Pelatihan dan subproses Pengujian. Subproses pelatihan digunakan untuk melatih model. Model yang terlatih kemudian diterapkan dalam subproses Pengujian. Kinerja model diukur selama fase Pengujian, berikut adalah tampilan dari tahapan cross validation:



Gambar 3.26 Proses di dalam cross validation

### 3.4.9. Apply model

*Apply Model* digunakan untuk menghubungkan model algoritma dengan *performance*, *apply model* dapat juga berperan sebagai penghubung. Menyetujui bahwa algoritma yang digunakan tersebut akan dikoneksikan ke fitur *performance*.



Gambar 3.27 Tahapan apply model



### 3.4.10. Performance

Operator ini digunakan untuk secara statistik mengevaluasi kekuatan dan kelemahan klasifikasi biner, setelah model yang terlatih telah diterapkan pada data berlabel.



Gambar 3.28 Tahapan Performance

### 3.5. Alur penelitian

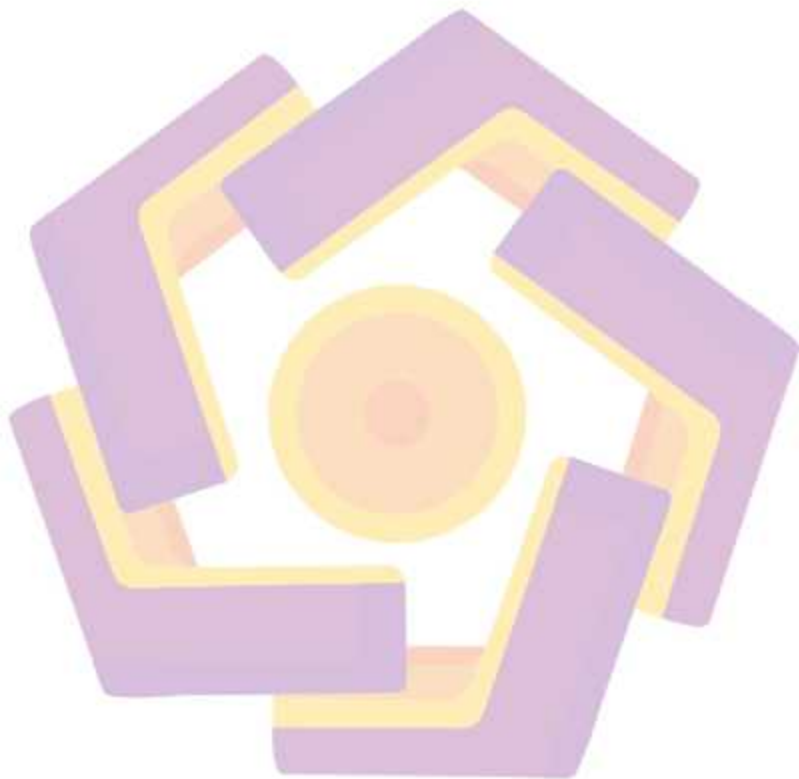
Pada penelitian ini alur atau langkah-langkah yang dilakukan dalam penelitian ini dijabarkan sesuai dengan gambar berikut:



Gambar 3.29 Alur Penelitian

Ada empat metode perbandingan yang akan digunakan yaitu klasifikasi dengan k-NN, klasifikasi dengan Naive Bayes, klasifikasi dengan k-NN dengan PSO, dan klasifikasi dengan Naive bayes dengan PSO. Pada tahapan perama dataset akan diambil melalui media sosial twitter. Kemudian dilanjutkan dengan preprocessing, yaitu pembersihan data, case folding, normalization, dan stemming. Kemudian

setelah preprocessing dilanjutkan dengan proses pelabelan data melalui sentiment scoring, dengan memberikan nilai positive, negative dan neutral pada setiap data. Proses terakhir adalah proses perhitungan akurasi dari algoritma yang ada dan hasil akurasi akan dibandingkan untuk mengetahui algoritma terbaik.



## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

Metode training data ini dikembangkan dengan sentiment analisis dengan metode k-NN dan Naive Bayes dengan membuat label dari masing-masing baris data dengan metode statistikal learning dengan mengacu pada hasil dari sentiment analisis yang dilakukan dengan menggunakan Rapidminer Studio.

#### 4.1. Crawling data Twitter

Tabel 4.1 Contoh hasil crawling data twitter

Bismillahirramanirrahim. Semoga Allah SWT merestui dan meridhoi perjuangan kita semua. Aamiin. <a href="https://t.co/whlwgr0VGE">https://t.co/whlwgr0VGE</a>
Prabowo-Sandi Diprediksi Keruk Suara 'Silent Voters' <a href="https://t.co/T8u1zUJZ0u">https://t.co/T8u1zUJZ0u</a>
Are you ready, Pak @prabowo? Bismillah untuk besok. <a href="https://t.co/iZYe12whJ5">https://t.co/iZYe12whJ5</a>
RT @prabowo: Bismillahirramanirrahim. Semoga Allah SWT merestui dan meridhoi perjuangan kita semua. Aamiin. <a href="https://t.co/whlwgr0VGE">https://t.co/whlwgr0VGE</a>
RT @maulinaantika02: Kpd Teman2 relawan Yg Ada di CIREBON @prabowo @sandiuono di Cirebon Timur tdk mempunyai Saksi dalam di TPS Saksi yg sdh...

Penelitian sentiment analisis sangat berkaitan dengan jumlah data untuk mendapatkan hasil yang terbaik. Penelitian ini dimulai dengan proses mengambil data tweet dengan query jokowi dan prabowo dengan jumlah masing-masing 2000 data.



Gambar 4.1 Grafik perbandingan crawling data

Dari tweet yang telah di ambil akan dijadikan sebagai dataset kemudian dilakukan proses seleksi data, agar data bervariasi dan tidak terdapat duplikasi data.

#### 4.2. Seleksi data

Tabel 4.2 Contoh hasil seleksi data

Dinamika di Detik-detik Terakhir Menjelang Pemilu 2019: Mampukah Kalahkan Jokowi?
YAH KOK JOKOWI GEMINI
@fkokr @jokowi lya itu dah...kok gak punya sisi kemanusiaan sama sekali kesanya
@Tungulsitomp1 @jokowi Sisi kemanusiaan dimana coba spa tau itu bapak d kasih baju itu, dan punyaanya cuman itu, seharusnya mau make baju 01/02 juga gak sepatasnya kayak gitu
@rgumelarp @jokowi Via bikini bottom

Proses seleksi data yaitu proses penghapusan duplikasi data untuk mendapatkan data yang lebih bervariasi. Seleksi data dilakukan dengan proses penghapusan duplikasi melalui penghapusan url, dan hastag.



Gambar 4.2 Hasil seleksi data

#### 4.3. Preprocessing

Untuk meningkatkan akurasi serta kecepatan proses sentiment dibutuhkan langkah preprocessing data. Preprocessing data dilakukan dengan melakukan penghapusan kata, angka dan simbol-simbol yang tidak berarti, mengubah kata yang tidak baku menjadi baku (Normalisasi) serta menghapus kata berimbuhan

menjadi kata dasar (Stemming). Melalui proses ini keceperan serta hasil klasifikasi data meningkat.

Tabel 4.3 Contoh hasil preprocessing data

dinamika di detik-detik terakhir jelang pemilu mampukah kalahkan jokowi
yah kok jokowi gemini
iya itu dah kok gak punya sisi kemanusiaan sama sekali kesanya
sisi kemanusiaan dimana coba sapa tau itu bapak kasih baju itu, dan punya hanya itu seharusnya mau pakai baju juga gak sepatasnya kayak begitu
via bikini bottom

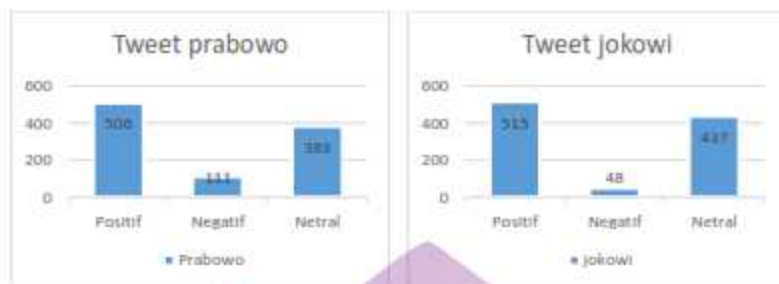
#### 4.4. Sentiment scoring

Dari hasil seleksi data yang diperoleh kemudian diambil 1000 data prabowo dan jokowi untuk diberi score sentiment. Sentiment scoring adalah proses pemberian nilai positif, negatif dan netral pada setiap kalimat dataset.

Tabel 4.4 Contoh hasil sentiment scoring

	A	B	C
1	Nomer	Score	Text
2	1	0	Belajar Untuk Berbakti, Raih Ilmu Untuk Bangsa
3	2	-2	Bilang Pak ANH [Agus Harimurti Yudhoyono, Kogasma Partai Demokrat, Iled], kita belajar dari koalisi,
4	3	-1	Garis keras
5	4	2	Kami tak bs lg mengajak & mengarahin umat utk memilih pasangan @prabowo @sandiwano di Pilpres 2019 nanti. Km takt berdosz jk meminta umat memilih pasangan yg jelas2 tdk diinginkan Allah. Krna Allah
6	5	-1	PILPRES,pilih presiden
7	6	-2	Prabowo lebih TNI daripada kebanyakan TNI.
8	7	-4	Saya pilih Prabowo karena janji-janji Jokowi banyak yang tidak terrealisasi.
9	8	-2	TIDAK MENGAMBIL GAJI, SERUPAH PUN
10	9	2	#Jakarta - Wakil Presiden Jusuf Kalla (JK) meminta para elite politik belajar dari pelukan Joko Widodo (Jokowi) dan Prabowo Subianto. Gausai debat capres-cawapres. Para elite harus menjadi

Proses scoring dilakukan dengan metode statistikal learning melalui pembobotan kata positif dan negatif yang menghasilkan nilai  $>0$  sebagai positif, dan  $<0$  sebagai negatif dan 0 sebagai netral.



Gambar 4.3 Grafik hasil pelabelan

#### 4.5. Klasifikasi sentiment analysis

Setelah data melalui proses preprocessing kemudian diberi label positif dan negatif kemudian dilakukan proses klasifikasi sentimen dengan menggunakan metode NB, KNN, PSO-KNN dan PSO-NB. Pada tahap ini dilakukan dua metode perhitungan yaitu metode perhitungan manual dan metode perhitungan otomatis. Metode perhitungan manual digunakan untuk mengetahui cara kerja algoritma secara mendetail sedangkan metode otomatis digunakan untuk mengetahui hasil penelitian secara menyeluruh.

##### 4.5.1. Perhitungan manual klasifikasi sentiment analysis

Perhitungan manual dilakukan dengan menggunakan data yang sudah mengalami preprocessing dan pembobotan TF-IDF. Ada 9 trining yaitu data ke-1 sampai dengan 9 dan 1 data uji yaitu data ke-10.

Tabel 4.5 Dataset setelah preprocessing

No	Text	Sentiment
1	emak anak jaditumbalpresiden jokowi widodo jokowi kecam a teror yang libat anak-anak anak-anak dalam lancar aksi	Negatif
2	lamun siro sekti ojo matenilamun siro banter ojo ndhisukilamun siro pintar ojo minteri ir jokowi widodo	Positif
3	kalau saya mah yang penting bukan sih jokowi widodo	Positif
4	tipe orang presiden nkri ir jokowi widodo sampai detik ini belum kategori orang-orang tidak jujur	Negatif
5	selalu bikin hati-hati adem tiap lihat bpk presiden jokowi bicara presiden jawab dialog ledi marina ada	Positif
6	ini kandung surat al fatihah arasemen di buat mei untuk bapak jokowi widodo moga sukses	Positif
7	rt semua orang-orang sudah tau kalau pak jokowi widodo undang syamsuri firdaus menang musabaqah tilawatil quran mtq internasional ke	Positif
8	sukses selalu bapak presiden jokowi widodo salami dari malu tengah desa walmusi c seram utara timur kabi prop maluku	Negatif
9	nyata sikap-sikap front juang rakyat hadap umum hasil mfu dan tndas fasis pemerintah jokowi presiden ir jokowi widodo pribadi tampil apa-apa ada kalem tidak mudah emotional tidak mudah marah dan sok	Negatif
10		???

Tabel 4.6 Dataset setelah pembobotan TF-IDF

	1	2	3	4	5	...	10
adem					0,242336	...	
aksi	0,104399					...	
al						...	
anak	0,821993					...	
...	...	...	...	...	...	...	...
Widodo	0,304399	0,188982	0,3	0,235702		...	0,267

#### 4.5.1.1. Perhitungan manual klasifikasi sentiment analisis dengan k-NN

Perhitungan KNN secara sederhana dapat dilihat pada flowchart di bawah

ini:



Gambar 4.4 flowchart KNN

Asumsikan data ke-10 adalah data yang ingin kita ketahui nilainya. Dari flowchart diatas dapat dilakukan perhitungan dengan proses sebagai berikut:

#### 7. Menentukan dataset sentiment

Langkah pertama adalah memasukan dataset pelatihan yang telah mengalami preprocessing dan pembobotan pada tabel 4.6.

#### 8. Menentukan jarak euclidean

Langkah selanjutnya adalah perhitungan jarak euclidean dengan rumus sebagai berikut:

$$dist(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (10)$$

Dari rumus diatas didapatkan hasil sebagai berikut :

Tabel 4.7 Perhitungan jarak euclidean

	1	2	3	4	5	6	7	8	9	10
adem					0,059					
aka	0,027									
el						0,100				
anak	0,070									
widodo	0,011	0,006	0,054	0,001		0,002	0,001	0,000		
Jarak euclidean	0,938	0,955	0,780	0,884	0,808	0,897	0,940	0,880	0,906	0,000

#### 9. Menentukan k tetangga terdekat

Kemudian langkah selanjutnya adalah menentukan k tetangga terdekat dari data ke-10. Yaitu nilai jarak terkecil dari setiap kalimat yang ada.

Tabel 4.8 Perhitungan k tetangga

	1	2	3	4	5	6	7	8	9
Sentiment	N	P	P	N	P	P	P	N	N
Euclidean	0,375	0,382	0,312	0,354	0,347	0,359	0,370	0,355	0,380
K tetangga	0	8	1	3	2	5	7	4	9

Dari hasil diatas data testing dapat kita prediksi dengan nilai k sebagai berikut.

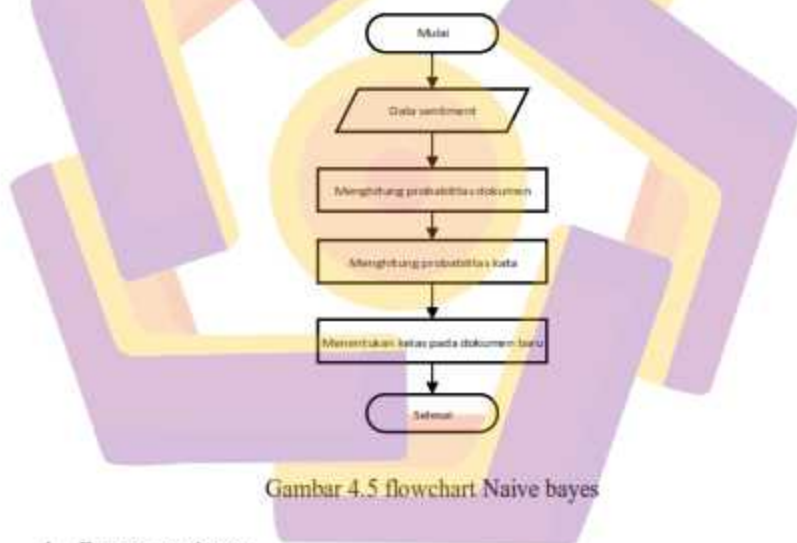


Tabel 4.9 Prediksi data testing berdasarkan nilai k

	Jumlah Positif	Jumlah Negatif	Prediksi
K=1	1	0	Positif
K=2	2	0	Positif
K=3	2	1	Positif
K=4	2	2	Netral
K=5	3	2	Positif

#### 4.5.1.2. Perhitungan manual klasifikasi sentiment analisis dengan Naive bayes

Proses klasifikasi sentiment dengan menggunakan naive bayes dilakukan dengan beberapa langkah sebagai berikut :



Gambar 4.5 flowchart Naive bayes

##### 1. Dataset sentiment

Langkah pertama menentukan bobot dari setiap term dataset sentiment pada tabel 4.5 dengan menggunakan TF-IDF seperti terlihat pada tabel 4.6.

##### 2. Menghitung probabilitas dokument

Langkah selanjutnya adalah menghitung banyaknya kemunculan dokumen positif dan negatif pada dataset dengan rumus sebagai berikut.

$$P(v_j) = \frac{P(v_j)}{\sum P} \dots \dots \dots (11)$$

Dari rumus diatas didapatkan probabilitas dokumen sebagai berikut:

Tabel 4.10 Perhitungan probabilitas dokumen

	Jumlah	Probabilitas
Positif	6	0,6
Negatif	4	0,4

3. Menghitung probabilitas kata pada setiap dokument

Langkah selanjutnya menghitung probabilitas kata akan muncul dalam kelas (data trining) positif dan negatif menggunakan rumus :

$$P(w_k|v_j) = \frac{n_k+1}{n+|\text{Vocabulary}|} \dots \dots \dots (12)$$

Dengan menggunakan rumus diatas menghasilkan data sebagai berikut :

Tabel 4.11 Perhitungan probabilitas kata

	Total kejadian	Positif	Negatif	P(kata  Positif)	P(kata  Negatif)
adem	0,243	0,243	0,000	0,025	0,023
aksi	0,164	0,000	0,164	0,021	0,027
ai	0,316	0,316	0,000	0,027	0,023
anak	0,822	0,000	0,822	0,021	0,042
...	...	...	...	...	...
widodo	2,182	1,515	0,667	0,052	0,039

4. Menentukan kelas pada dokument baru

Proses diatas adalah proses pembelajaran dokumen pada tiap kelas, dengan menggunakan hasil belajar tersebut untuk melakukan klasifikasi pada dokument baru dengan rumus :

$$v_{NB} = \frac{\arg \max_{v_j \in V} P(v_j)}{v_j \in V} \parallel P(a_i|v_j) \dots \dots \dots (13)$$

Dengan menggunakan rumus tersebut dihasilkan data sebagai berikut :

Text		
Pilih orang pintar jokowi widodo		

Kata	P(Kata Negatif)	P(Kata Positif)
pilih	0,005617978	0,005154639
pilih	0,021	0,023
orang	0,030	0,040
pintar	0,024	0,023
jokowi	0,057	0,049
widodo	0,052	0,039

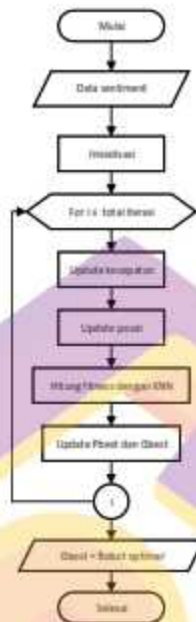
Total	
$P(\text{Positif}) * P(\text{kata} \text{Positif})$	0,0000000269
$P(\text{Negatif}) * P(\text{kata} \text{Negatif})$	0,0000000162

Gambar 4.6 Penentuan kelas dokumen testing

Karena jumlah positif lebih tinggi dari pada negatif maka kesimpulanya kata tersebut bernilai : Positif

#### 4.5.1.3. Perhitungan manual klasifikasi sentiment analisis metode k-NN dan PSO

Proses optimasi K-NN menggunakan algoritma PSO ditunjukkan pada gambar di berikut :



Gambar 4.7 flowchart KNN-PSO

Proses PSO-KNN dilakukan dengan beberapa langkah sebagai berikut :

1. Data sentiment

Langkah pertama adalah memasukan dataset pelatihan yang telah mengalami preprocessing dan pembobotan seperti yang terlihat pada tabel 4.5 dan 4.6.

2. Inisialisasi

a. Inisialisasi kecepatan awal. Pada iterasi ke-0 dapat dipastikan bahwa kecepatan awal semua partikel adalah 0.

b. Inisialisasi Pbest dan Gbest. Pada iterasi ke-0, Pbest dan Gbest akan disamakan dengan nilai posisi awal, sedangkan Gbest akan akan dipilih dari salah satu Pbest dengan nilai fitness tertinggi.

c. Inisialisasi  $c_1=c_2=1$ ,  $W_{min} = 0,4$  dan  $W_{max} = 0,9$  dan  $N=100$ .

### 3. Update kecepatan.

Untuk melakukan update kecepatan digunakan rumus berikut :

$$v_{i,j}^{t+1} = w \cdot v_{i,j}^t + c_1 \cdot r_1 (P_{best\ i,j} - x_{i,j}^t) + c_2 \cdot r_2 (G_{best} - x_{i,j}^t) \dots \dots \dots (14)$$

Tabel 4.12 Hasil perhitungan update kecepatan iteasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					-0,146					
aksi	-0,099									
al						-0,190				
anak	-0,453									
...	...	...	...	...	...	...	...	...	...	...
widodo	-0,099	-0,113	-0,300	-0,141		-0,190	-0,146	-0,180		-0,160

### 4. Update posisi.

Untuk melakukan update kecepatan digunakan rumus sebagai berikut :

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \dots \dots \dots (15)$$

Tabel 4.13 Hasil perhitungan update posisi iteasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					0,097					
aksi	0,066									
al						0,126				
anak	0,323									
...	...	...	...	...	...	...	...	...	...	...
widodo	0,066	0,076	0,200	0,094		0,126	0,097	0,107		0,107

### 5. Hitung fitness dengan KNN

Asumsikan data ke-10 adalah data yang ingin kita ketahui nilainya. Untuk menghitung fitness dengan KNN, dilakukan perhitungan jarak euclidean dengan rumus sebagai berikut:

$$dist(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (16)$$

Kemudian langkah selanjutnya adalah menentukan k tetangga terdekat dari data ke-10. Yaitu nilai jarak terkecil dari setiap kalimat yang ada.

Tabel 4.14 Hasil penentuan k tetangga

	1	2	3	4	5	6	7	8	9
Sentiment	N	P	P	N	P	P	P	N	N
dist(x,y)	0,375	0,382	0,312	0,354	0,347	0,359	0,370	0,355	0,380
K tetangga	0	8	1	3	2	5	7	4	9

Dengan menggunakan nilai  $k=5$  dihasilkan nilai positif (P) = 3 dan Negatif (N) = 2. Dengan demikian pada iterasi ke-1 prediksi data ke-10 adalah sentiment positif.

#### 6. Update Pbest dan Gbest

Setelah nilai fitness ditemukan langkah Selanjutnya adalah menentukan nilai terbaik lokal yang merupakan partikel terbaik lokal yang disebut sebagai  $P_{best,j}$  serta nilai terbaik global yang merupakan posisi terbaik dari seluruh kawanan yang disebut sebagai  $G_{best,j}$ . Dengan membandingkan nilai fitness dari Pbest pada iterasi sebelumnya dengan fitness dari update Posisi. Nilai yang terbaik akan menjadi Pbest yang baru pada iterasi selanjutnya. Dari hasil perhitungan ditemukan nilai  $G_{best,j} = 0,329$  dan Pbest sebagai berikut.

Tabel 4.15 Update Pbest iterasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					0,097					
aksi	0,066									
af					0,126					
anak	0,329									
...	...	...	...	...	...	...	...	...	...	...
widodo	0,066	0,076	0,200	0,094		0,126	0,097	0,107		0,107

#### 7. Ulangi langkah ke-3 sampai ke-6 hingga mencapai iterasi maksimum.

Dari 4 iterasi yang dilakukan didapatkan jarak euclidean terhadap data ke-10 sebagai berikut:

Tabel 4.16 Jarak euclidean setiap iterasi

	D 1	D 2	D 3	D 4	D 5	D 6	D 7	D 8	D 9
iterasi 1	0,375	0,382	0,312	0,354	0,347	0,359	0,376	0,355	0,386
iterasi 2	0,411	0,400	0,128	0,305	0,391	0,326	0,451	0,424	0,417
iterasi 3	0,380	0,351	0,325	0,346	0,414	0,428	0,439	0,448	0,453
iterasi 4	0,402	0,394	0,324	0,353	0,433	0,426	0,467	0,450	0,469

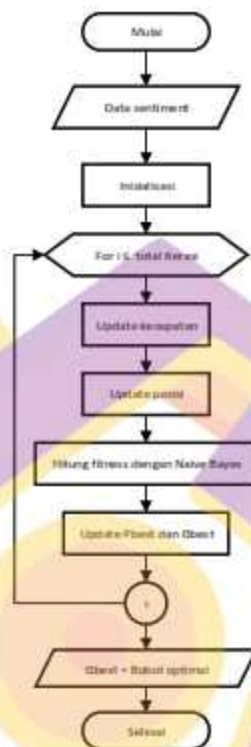
Agar lebih mudah membaca pergerakan jarak, data di atas dibuat menjadi diagram seperti pada gambar dibawah :



Gambar 4.8 Diagram pergerakan jarak euclidean setiap iterasi

#### 4.5.1.4. Perhitungan manual klasifikasi sentiment analisis metode Naive Bayes dan PSO

Proses klasifikasi sentiment dilakukan dengan beberapa langkah sebagai berikut :



Gambar 4.9 Flowchart PSO-NB

### 1. Data sentiment

Langkah pertama adalah memasukan dataset pelatihan yang telah mengalami preprocessing dan pembobotan seperti yang terlihat pada tabel 4.5 dan 4.6.

### 2. Inisialisasi

- a. Inisialisasi kecepatan awal. Pada iterasi ke-0 dapat dipastikan bahwa kecepatan awal semua partikel adalah 0.



- b. Inisialisasi Pbest dan Gbest. Pada iterasi ke-0, Pbest dan Gbest akan disamakan dengan nilai posisi awal, sedangkan Gbest akan akan dipilih dari salah satu Pbest dengan nilai fitness tertinggi.
- d. Inisialisasi  $c_1=c_2=1$ ,  $W_{min} = 0,4$  dan  $W_{max} = 0,9$  dan  $N=100$ .

### 3. Update kecepatan.

Untuk melakukan update kecepatan digunakan rumus berikut :

$$v_{i,j}^{t+1} = w \cdot v_{i,j}^t + c_1 \cdot r_1 (P_{best\ i,j} - x_{i,j}^t) + c_2 \cdot r_2 (G_{best} - x_{i,j}^t) \dots \dots \dots (17)$$

Tabel 4.17 Hasil perhitungan update kecepatan iteasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					0,348					
eksi	0,395									
al						0,303				
anak	0,000									
---	---	---	---	---	---	---	---	---	---	---
widodo	0,395	0,380	0,193	0,352		0,303	0,348	0,333		0,333

### 4. Update posisi.

Untuk melakukan update kecepatan digunakan rumus sebagai berikut :

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \dots \dots \dots (18)$$

Tabel 4.18 Hasil perhitungan update posisi iteasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					0,590					
eksi	0,559									
al						0,620				
anak	0,822									
---	---	---	---	---	---	---	---	---	---	---
widodo	0,559	0,569	0,693	0,587		0,620	0,590	0,600		0,600

### 5. Hitung fitness dengan Naive bayes

Langkah selanjutnya adalah perhitungan nilai fitness dengan Naive bayes.

Dengan cara menghitung probabilitas kata akan muncul dalam kelas positif dan

negatif menggunakan rumus :

$$P(w_k|v_j) = \frac{n_k+1}{n+|\text{Vocabulary}|} \dots \dots \dots (19)$$

Dengan menggunakan rumus diatas menghasilkan data sebagai berikut :

Tabel 4.19 Hasil perhitungan probabilitas kata

Data ke	Total kejadian	Jml positif	Jml negatif	p(kata positif)	p(kata negatif)
adem	0,590	0,590	0,000	0,015	0,011
aksi	0,559	0,000	0,559	0,010	0,017
al	0,620	0,620	0,000	0,016	0,011
anak	0,822	0,000	0,822	0,010	0,020
...	...	...	...	...	...
widodo	4,819	3,072	1,747	0,039	0,029

Proses diatas adalah proses pembelajaran dokumen pada tiap kelas, dengan menggunakan hasil belajar tersebut untuk melakukan klasifikasi pada dokumen baru dengan rumus :

$$v_{NB} = \frac{\arg \max_{v_j \in V} P(v_j) \prod P(a_i|v_j)}{\dots \dots \dots} (20)$$

Dengan menggunakan rumus tersebut dihasilkan data sebagai berikut :

Text	
Pilih orang pintar jokowi widodo	

Kata	P(kata Negatif)	P(kata Positif)
pilih	0,010	0,011
orang	0,016	0,019
pintar	0,015	0,011
jokowi	0,045	0,037
widodo	0,039	0,029

P(Negatif) * P(kata Negatif)	0,0000000026
P(Positif) * P(kata Positif)	0,0000000009

Gambar 4.10 Hasil klasifikasi dokumen baru

Karena jumlah positif lebih tinggi dari pada negatif maka kesimpulanya kata tersebut bernilai : Positif

#### 6. Update Pbest dan Gbest

Setelah nilai fitness ditemukan langkah Selanjutnya adalah menentukan nilai terbaik lokal yang merupakan partikel terbaik lokal yang disebut sebagai  $P_{best,j}$  serta nilai terbaik global yang merupakan posisi terbaik dari seluruh kawanan yang disebut sebagai  $G_{best,j}$ . Dengan membandingkan nilai fitness dari Pbest pada iterasi sebelumnya dengan fitness dari update Posisi. Nilai yang terbaik akan menjadi Pbest yang baru pada iterasi selanjutnya. Dari hasil perhitungan ditemukan nilai  $G_{best,j} = 0,822$  dan Pbest sebagai berikut.

Tabel 4.20 Pbest iterasi ke-1

	1	2	3	4	5	6	7	8	9	10
adem					0,590					
aki	0,559									
ai						0,620				
arak	0,822									
...	...	...	...	...	...	...	...	...	...	...
widodo	0,559	0,509	0,693	0,587		0,620	0,590	0,600		0,600

#### 8. Ulangi langkah ke-3 sampai ke-6 hingga mencapai iterasi maksimum.

Dari 4 iterasi yang dilakukan didapatkan nilai probabilitas pada dokument baru sebagai berikut:

Tabel 4.21 nilai probabilitas dokumen baru

	kata	pilih	orang	pintar	jokowi	widodo	jumlah
Iterasi 1	positif	0,010	0,016	0,015	0,045	0,039	0,0000000020
	negatif	0,011	0,019	0,011	0,037	0,029	0,0000000009
Iterasi 2	positif	0,010	0,018	0,019	0,064	0,055	0,0000000071
	negatif	0,011	0,020	0,011	0,051	0,041	0,0000000019
Iterasi 3	positif	0,010	0,020	0,021	0,077	0,065	0,0000000124
	negatif	0,011	0,021	0,011	0,061	0,049	0,0000000029
Iterasi 4	positif	0,010	0,021	0,019	0,087	0,057	0,0000000087
	negatif	0,011	0,023	0,011	0,051	0,041	0,0000000022

Agar lebih mudah membaca pergerakan probabilitas, data di atas dibuat menjadi diagram seperti pada gambar dibawah :



Gambar 4.11 Diagram probabilitas kata dokumen baru

#### 4.5.2. Klasifikasi sentiment anllsils dengan Rapidminer

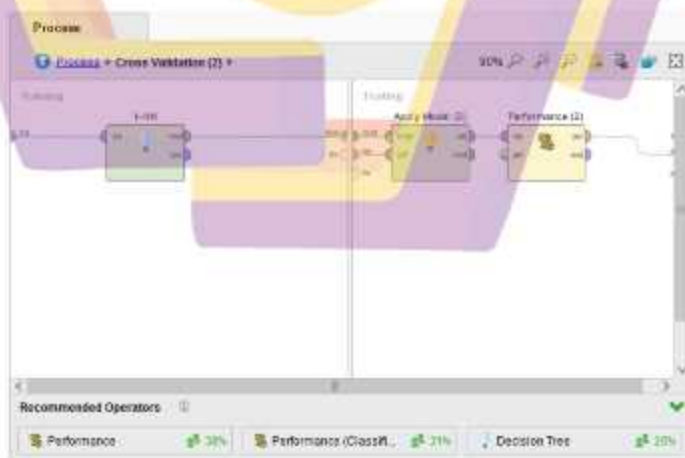
Pada perhitungan klasifikasi ini dilakukan dengan menggunakan tools Rapidminer. Perhitungan pada tahap ini menggunakan 563 data jokowi dan 617 data prabowo. Langkah-langkah proses dapat dilihat pada gambar dibawah ini.



Gambar 4.12 Klasifikasi sentiment analisis

#### 4.5.2.1. Perhitungan Rapidminer klasifikasi sentiment analisis dengan k-NN

Proses klasifikasi sentiment berada di dalam proses cross validation yang terdapat pada gambar diatas. Metode k-NN diterapkan pada langkah ini dan dapat dilihat pada gambar berikut :



Gambar 4.13 Proses sentimen analisis menggunakan metode k-NN

Pada penelitian ini penulis melakukan beberapa percobaan untuk mendapatkan nilai akurasi tertinggi. Adapun percobaan yang dilakukan adalah mengubah nilai k dari 1 sampai dengan 10. Perhitungan pertama menggunakan tweet jokowi dan dihasilkan akurasi dan kecepatan proses sebagai berikut.

Tabel 4.22 Hasil klasifikasi sentiment jokowi dengan k-NN

Nilai k	Hasil akurasi (%)	Waktu (detik)
1	97,69	3
2	97,69	3
3	97,51	3
4	96,80	2
5	96,45	3
6	96,31	3
7	96,10	3
8	96,10	2
9	95,56	2
10	95,74	2

Penelitian selanjutnya dengan menggunakan tweet prabowo dihasilkan tingkat akurasi sebagai berikut:

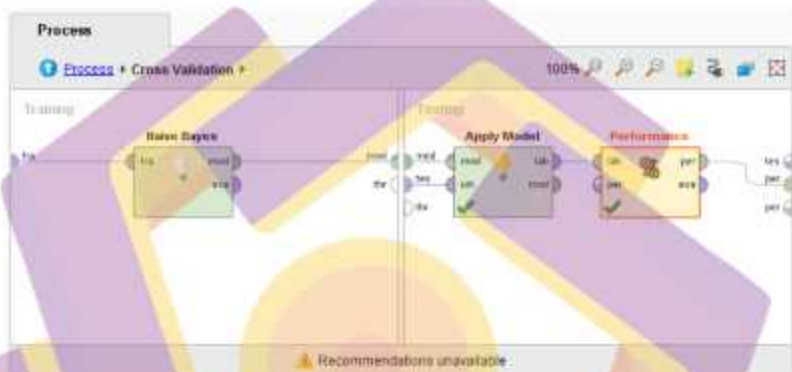
Tabel 4.23 Hasil klasifikasi sentiment prabowo dengan k-NN

Nilai k	Hasil akurasi (%)	Waktu (detik)
1	93,35	5
2	93,51	4
3	92,38	5
4	92,86	5
5	92,21	4
6	92,05	4
7	89,62	4
8	89,62	4
9	89,30	5
10	89,13	4

Dari hasil perhitungan diperoleh nilai k terbaik adalah k=2, Kemudian nilai k=2 ini akan digunakan untuk perhitungan PSO-KNN selanjutnya. Selain itu hasil akurasi tertinggi yang diperoleh akan digunakan untuk perbandingan antara metode satu dan lainnya.

#### 4.5.2.2. Perhitungan Rapidminer klasifikasi sentiment analisis dengan Naive Bayes

Proses klasifikasi sentiment berada di dalam proses cross validation yang terdapat pada gambar diatas. Metode Naive bayes diterapkan pada langkah ini dan dapat dilihat pada gambar berikut.



Gambar 4.14 Klasifikasi sentimen dengan metode Naive bayes

Dari hasil penelitian dihasilkan tingkat akurasi sebagai berikut:

Tabel 4.24 Hasil klasifikasi sentiment jokowi dengan Naive bayes

Hasil akurasi (%)	Waktu (Detik)
96,09	3

Penelitian selanjutnya dengan menggunakan tweet prabowo dihasilkan tingkat akurasi sebagai berikut:

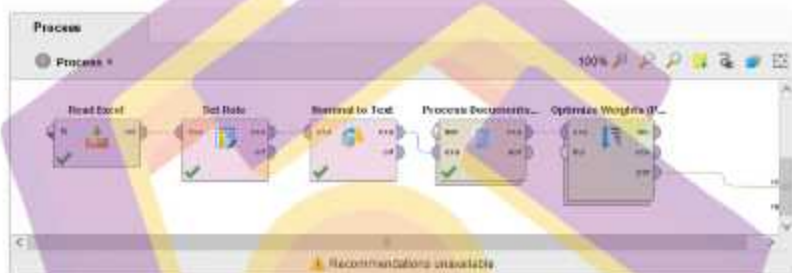
Tabel 4.25 Hasil klasifikasi sentiment jokowi dengan Naive bayes

Hasil akurasi (%)	Waktu (detik)
91,90	3

Dari perhitungan di atas didapatkan hasil akurasi 96,09% untuk jokowi dan 91,90% untuk prabowo, kemudian dari hasil perhitungan ini akan digunakan untuk perbandingan tingkat akurasi dengan metode lainnya.

#### 4.5.2.3. Perhitungan Rapidminer klasifikasi sentiment analisis dengan PSO dan k-NN

Optimize weight (PSO) digunakan untuk mengoptimasi peningkatan akurasi pada sentiment jokowi dan prabowo. Meskipun waktu yang dibutuhkan untuk proses klasifikasi semakin tinggi, namun dalam penelitian ini berhasil memperlihatkan peningkatan akurasi pada metode k-NN dan Naive bayes.



Gambar 4.15 Proses sentimen analisis menggunakan metode k-NN dan PSO

Perbedaan proses klasifikasi sentiment dengan menggunakan k-NN dengan optimasi PSO adalah dengan menambahkan tools Optimize Weights dan memasukan cross validation pada tools tersebut.

Pada perhitungan ini penulis melakukan beberapa percobaan untuk menghasilkan akurasi terbaik yaitu dengan mengubah jumlah partikel dari setiap populasi serta mengubah jumlah iterasi. Dari hasil penelitian dihasilkan tabel sebagai berikut:



Tabel 4.26 Hasil klasifikasi sentiment dengan k-NN - PSO

Jokowi PSO-KNN				Prabowo PSO-KNN			
Partikel	Maksimal Iterasi	Hasil akurasi (%)	Waktu (Menit:Detik)	Partikel	Maksimal Iterasi	Hasil akurasi (%)	Waktu (Menit:Detik)
10	10	98,41	1;19	10	10	94,98	5;13
	20	98,58	2;37		20	95,79	9;50
	30	98,76	3;54		30	96,26	13;20
	40	98,76	5;02		40	95,62	19;46
	50	98,76	6;21		50	95,79	21;19
20	10	98,58	2;44	20	10	95,13	10;27
	20	98,93	4;58		20	96,11	19;29
	30	98,93	6;20		30	96,60	28;08
	40	98,94	8;24		40	96,43	37;57
	50	98,94	11;34		50	96,45	44;46
30	10	98,40	4;13	30	10	95,33	10;31
	20	98,94	6;52		20	95,65	30;44
	30	98,76	11;22		30	96,10	44;15
	40	98,59	14;35		40	96,21	58;03
	50	98,77	16;56		50	96,43	01;12;33
40	10	98,40	6;01	40	10	94,49	17;11
	20	98,59	14;31		20	95,55	32;15
	30	98,75	18;37		30	96,34	40;33
	40	98,76	22;52		40	96,01	01;01;35
	50	98,76	27;26		50	96,93	01;16;42
50	10	98,93	6;58	50	10	94,67	18;41
	20	98,93	12;35		20	95,21	34;21
	30	98,94	17;56		30	95,95	50;10
	40	98,94	38;41		40	96,60	01;07;09
	50	98,94	47;07		50	96,60	01;24;26

Dari data di atas dapat disimpulkan bahwa jumlah partikel dalam populasi dan jumlah iterasi mempengaruhi akurasi dan kecepatan perhitungan PSO-KNN. Dari data di atas didapatkan akurasi tertinggi didapatkan pada sentiment Jokowi adalah pada jumlah partikel 50 dan jumlah iterasi 30, yaitu sebanyak 98,94% dengan waktu 17,56 menit. Dan pada sentiment Prabowo pada jumlah partikel 40 dan jumlah iterasi 50, yaitu sebanyak 96,93% dengan waktu 01;16;42 menit.

#### 4.5.2.4. Perhitungan Rapidminer klasifikasi sentiment analisis dengan Naive Bayes dan PSO.

Pada perhitungan ini penulis melakukan beberapa percobaan yaitu dengan mengubah jumlah partikel dari setiap populasi dan juga mengubah jumlah iterasi. Dengan menggunakan tweet jokowi dihasilkan akurasi sebagai berikut:

Tabel 4.27 Hasil klasifikasi sentiment jokowi dengan PSO - NB

Jokowi PSO-NB				Prabowo PSO-NB			
Partikel	Iterasi	Hasil akurasi (%)	Waktu (Menit:Detik)	Partikel	Iterasi	Hasil akurasi (%)	Waktu (Menit:Detik)
10	10	97,34	0:50	10	10	93,20	1:32
	20	97,33	1:26		20	94,33	2:48
	30	98,06	1:37		30	94,80	3:50
	40	97,87	2:29		40	94,01	4:59
	50	98,04	3:11		50	95,13	6:06
20	10	97,34	1:42	20	10	94,65	2:51
	20	97,88	2:49		20	95,30	5:17
	30	98,06	3:55		30	95,15	7:24
	40	98,24	3:09		40	95,95	9:02
	50	98,05	6:21		50	96,11	11:29
30	10	97,52	1:32	30	10	94,01	5:00
	20	98,40	2:52		20	94,98	9:00
	30	98,70	4:27		30	94,98	12:04
	40	98,58	5:01		40	95,84	15:12
	50	98,23	9:25		50	96,05	18:35
40	10	97,87	1:33	40	10	94,65	5:47
	20	97,87	3:36		20	95,63	10:16
	30	98,05	6:41		30	95,78	15:44
	40	98,50	8:02		40	96,06	21:12
	50	98,38	9:23		50	96,01	26:40
50	10	97,87	3:22	50	10	94,98	7:16
	20	98,05	5:54		20	95,79	12:25
	30	98,23	7:54		30	95,72	17:34
	40	98,40	8:26		40	95,34	22:43
	50	98,41	13:34		50	95,51	27:42

Dari data di atas dapat disimpulkan bahwa jumlah partikel dalam populasi dan jumlah iterasi mempengaruhi akurasi dan kecepatan perhitungan PSO-NB. Dari data di atas didapatkan akurasi tertinggi didapatkan pada sentiment jokowi pada jumlah partikel 30 dan jumlah iterasi 30, yaitu sebanyak 98,76% dengan waktu 4,27

menit. Dan pada sentiment prabowo pada jumlah partikel 40 dan jumlah iterasi 40, yaitu sebanyak 96,06% dengan waktu 21,12 menit.

#### 4.5.5. Perbandingan klasifikasi sentiment

Setelah penelitian berhasil dilakukan dan mendapatkan hasil akurasi dengan menggunakan algoritma k-NN, NB, PSO-KNN, dan PSO-NB. Selanjutnya hasil akurasi yang didapatkan dibandingkan untuk dapat diketahui metode mana yang terbaik dari penelitian ini. Berikut adalah diagram hasil perbandingan:

Tabel 4.30 Perbandingan klasifikasi sentiment Jokowi

	k-NN	NB	PSO - k-NN	PSO - NB
Akurasi (%)	97,09	90,09	98,94	98,70
Waktu (Menit; detik)	0;3	0;2	17;56	4;27

Tabel 4.31 Perbandingan klasifikasi sentiment Prabowo

	k-NN	NB	k-NN + PSO	NB + PSO
Akurasi (%)	93,51	91,90	96,93	96,06
Waktu (Jam; Menit; Detik)	4	3	01:16:42	21,12

## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Berdasarkan hasil penelitian dan percobaan data tweet yang telah diolah menjadi sentiment analisis dengan menggunakan metode k-NN, Naive Bayes dan PSO dapat diambil beberapa kesimpulan yaitu:

- a. Pengujian menggunakan k-NN didapatkan hasil akurasi terbesar dengan nilai k=1 dan k=2 sebanyak 97,69% dengan waktu 3 detik pada sentiment jokowi, sedangkan pada sentiment prabowo pada nilai k=2 sebesar 93,51% dengan waktu 4 detik.
- b. Pengujian menggunakan Naive bayes didapatkan hasil akurasi 96,09% dengan waktu 2 detik untuk sentiment jokowi dan 91,90% dengan waktu 3 untuk sentiment prabowo.
- c. Pengujian menggunakan PSO-KNN didapatkan hasil akurasi terbesar dengan nilai k=2, pada partikel 50 dan iterasi 30 sebanyak 98,94% dengan waktu 17 menit 56 detik pada sentiment jokowi, sedangkan pada sentiment prabowo pada nilai partikel 40 dan iterasi 50 sebesar 96,93% dengan waktu 1 jam 16 menit 42 detik.
- d. Pengujian menggunakan PSO-NB didapatkan hasil akurasi terbesar dengan pada partikel 30 dan iterasi 30 sebanyak 98,76% dengan waktu 4 menit 27 detik pada sentiment jokowi, sedangkan pada sentiment prabowo pada nilai partikel 40 dan iterasi 40 sebesar 96,06% dengan waktu 21 menit 12 detik.

- e. Pengujian optimasi PSO pada algoritma k-NN dan Naive Bayes berhasil meningkatkan akurasi meskipun membutuhkan waktu proses cukup lama.

## 5.2. Saran

Berdasarkan penelitian yang telah dilakukan, penulis mengusulkan beberapa saran untuk pengembangan penelitian ini agar dapat ditingkatkan:

- a. Penelitian selanjutnya diharapkan dapat membandingkan algoritma lain supaya dapat mengetahui algoritma terbaik.
- b. Penelitian selanjutnya diharapkan mampu melakukan penelitian lebih mendalam pada algoritma PSO.



## DAFTAR PUSTAKA

### PUSTAKA BUKU

- Ade Djohar, M. (2012). *Metode Penelitian Teknik Informatika*. Yogyakarta: Deepublish .
- Bungin, B. (2005). *Metodologi Penelitian Kuantitatif*. Jakarta: Kencana.
- Cvijikj, I. P. (2013). *Online engagement factors on Facebook brand pages*. New York: Springer Vienna Publisher.
- Nuswaka, T. &. (2003). *Sentiment analysis: capturing favorability using natural*. New York: ACM.
- Sarwono, J. (2006). *Metode Penelitian Kuantitatif dan Kualitatif*. Yogyakarta: Graha Ilmu.
- Sismono, H. (2005). *Pengantar Logika Informatika*. Yogyakarta: Andi Offset.
- Suarga. (2006). *Algoritma Pemrograman*. Yogyakarta: Andi Offset.

### PUSTAKA MAKALAH, JURNAL ILMIAH ATAU PROSIDING

- Aliandu, P. (2013). Twitter Used by Indonesian President : An Sentiment Analysis of Timenline. *ISICO*, 713-717.
- Antonius Rachmat C, Y. L. (2016). Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan *Naive Bayes*. *JUISI*, 2(2), 1-9.
- Defri Rosdiyansyah, S. A. (t.thn.). Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicon.

- Dubey, S. A. (2017). Clustering and Sentiment Analysis on Twitter Data. *International Conference on Telecommunication and Networks (TEL-NET)*, 5.
- Husin Muhamad, C. A. (2017). OPTIMASI *NAIVE BAYES CLASSIFIER* DENGAN MENGGUNAKAN *PARTICLE SWARM OPTIMIZATION* PADA DATA IRIS. *JTIK*, 4(3), 180-184.
- Joda Pahlawan Romadhona Tanjung, M. A. (2017). Klasifikasi *Tweets* pada Twitter dengan menggunakan metode fuzzy K-Nearest Neighbour (Fuzzy K-NN) dan Query Expansion Berbasis Apriori. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(5), 405-414.
- Manongga, A. M. (2018). Sentimen Analisis *Tweet* Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan Naive Bayes. *SIMETRIS*, 313-318.
- Mohammad Zoqi Sarwani, W. F. (2015). Analisis Twitter untuk mengetahui karakter seseorang menggunakan algoritma *Naive Bayes classifier*. *SNSINDO*, 7.
- Novichasari, S. I. (2017). *Particle Swarm Optimization* Meningkatkan Akurasi *Naive Bayes Classifier*. *UDINUS*, 1-5.
- Nur Hayatin, M. M. (2014). Opinion Extraction of Public Figure Based on Sentiment Analysis in Twitter. *IPTEK Journal of Engineering*, 1(1), 9-14.
- Robert Habibi, D. B. (2016). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. *INFORMATIKA*, 12(1), 103-109.

- Syaputra, R. &. (2016). Analisis Sentimen Pada Sosial Media Twitter Terhadap Politik Di Indonesia Menggunakan Text Mining Dengan Metode *Naive Bayes Classifier*. *IJCCS*, 1-8.
- Taufik Hidayatulloh, A. H. (2016). Klasifikasi Sel Tunggal PAP Smear Berdasarkan Analisis Fitur Berbasis *Naive Bayes Classifier* dan *Particle Swarm Optimization*. *SWABUMI*, 186-193.

#### **PUSTAKA LAPORAN PENELITIAN**

- Deviyanto, A. (2018). *Sentimen Analisis Tweet Pornografi Kaum Homoseksual Sentimen Analisis Tweet Pornografi Kaum Homoseksual Sentimen Analisis Tweet Pornografi Kaum Homoseksual*. Yogyakarta: UIN Kalijaga.
- Hasibuan, Z. A. (2007). *Metodologi Penelitian pada bidang ilmu komputer dan teknologi Informasi*. Jakarta: Fakultas Ilmu Komputer Universitas Indonesia
- Rasyadi, M. H. (2017). *Analisis Sentimen Pada Twitter Menggunakan Metode Naive Bayes (Studi Kasus Pemilihan Gubernur DKI Jakarta 2017)*. Bogor: Institut Pertanian Bogor.
- Rizal, M. (2017). *Analisis Sentimen Pengguna Twitter Terhadap Objek Pariwisata Di Indonesia Menggunakan Algoritma Pengolahan Deep Natural Language Dari Ibm Insights Untuk Twitter*. Makasar: UIN Alauddin Makasar.



**PUSTAKA ELEKTRONIK**

Giummole, F. O. (2018, November 10). *A Study on Microblog and Search Engine*

*User Behaviors*. Diambil kembali dari How Twitter Trending Topics Help

Predict Google Hot Queries in HUMAN:

[https://iris.unive.it/handle/10278/38460#\\_WiKiAWiCzIU](https://iris.unive.it/handle/10278/38460#_WiKiAWiCzIU) (

