

**SKRIPSI**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP PESERTA  
PEMILU PADA MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
PERBANDINGAN METODE *NAIVE BAYES CLASSIFIER* DAN  
*SUPPORT VECTOR MACHINE***

Disusun Oleh :

Fany Syafira Nurul Kilah

0651 15 125



**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN  
ALAM  
UNIVERSITAS PAKUAN  
BOGOR  
2019**



## HALAMAN PENGESAHAN

Judul : Analisis Sentimen Masyarakat Terhadap Peserta Pemilu Pada Media Sosial *Twitter* Menggunakan Metode *Naive Bayes Classifier* Dan *Support Vector Machine*  
Nama : Fany Syafira Nurul Kilah  
NPM 065115125

**Mengesahkan,**

Pembimbing II  
Program Studi Ilmu Komputer FMIPA  
- UNPAK

Pembimbing I  
Program Studi Ilmu Komputer  
FMIPA - UNPAK

**Irma Anggraeni, M.Kom.**

**Dr. Prihastuti Harsani, M.Si.**

**Mengetahui,**

Ketua Program Studi Ilmu Komputer  
FMIPA - UNPAK

Dekan  
FMIPA - UNPAK

**Dr. Prihastuti Harsani, M.Si.**

**Dr. Prasetyorini, M.Si**

## **PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI**

Dengan ini saya menyatakan bahwa:

Sejauh yang saya ketahui, karya tulis ini bukan merupakan karya tulis yang pernah di publikasikan atau sudah pernah dipakai untuk mendapatkan gelar sarjana di Universitas lain, kecuali pada bagian-bagian di mana sumber informasinya dicantumkan dengan cara referensi yang semestinya.

Demikian pernyataan ini saya buat dengan sebenar-benarnya. Apabila kelak dikemudian hari terdapat gugatan, penulis bersedia dikenakan sanksi sesuai dengan peraturan yang berlaku.

Bogor, 5 Agustus 2019

Fany Syafira Nurul Kilah

0651 15 125

## **PERNYATAAN PELIMPAHAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

Saya yang bertandatangan di bawah ini :

**Nama** : Fany Syafira Nurul Kilah  
**NPM** : 065115125  
**Judul Skripsi** : Analisis Sentimen Masyarakat Terhadap Peserta Pemilu  
Pada Media Sosial *Twitter* Menggunakan Perbandingan  
Metode *Naive Bayes Classifier* Dan *Support Vector  
Machine*

Dengan ini saya menyatakan bahwa Paten dan Hak Cipta dari produk Skripsi dan Tugas Akhir di atas adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apapun kepada perguruan tinggi manapun.

Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan Paten, Hak Cipta dari karya tulis saya kepada Universitas Pakuan.

Bogor, 5 Agustus 2019

Fany Syafira Nurul Kilah

0651 15 125

## RIWAYAT HIDUP



Fany Syafira Nurul Kilah tinggal di Kota Sukabumi. Lahir di Jampangkulon Sukabumi pada tanggal 15 September 1996 sebagai anak pertama dari tiga bersaudara. Penulis memulai pendidikan di Sekolah Dasar yang bertempat di SDN 1 Bojonggenteng, kemudian tahun 2009. Masuk SMPN 1 Jampangkulon dan penulis adalah Alumni dari SMAN 1 Jampangkulon di Jampangkulon Sukabumi. Pada tahun 2015 penulis meneruskan pendidikan ke Universitas Pakuan Bogor, Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam. Selama di Universitas Pakuan, penulis pernah aktif di Himpunan Mahasiswa Komputer (HIMAKOM) sebagai sekretaris periode 2017 dan sebagai Asisten Praktikum Labkom angkatan 2015. Pada bulan Juli tahun 2019 penulis menyelesaikan penelitian dengan judul “Analisis Sentimen Masyarakat Terhadap Peserta Pemilu Pada Media Sosial *Twitter* Dengan Perbandingan Metode *Naive Bayes Classifier* Dan *Support Vector Machine*”.

## RINGKASAN

Dalam penelitian ini akan dilakukan analisis sentimen masyarakat terhadap peserta pemilu 2019 yang diungkapkan melalui jejaring sosial Twitter. Ada beberapa tahap untuk melakukan analisis sentimen, diantaranya adalah tahap pengumpulan data (*scraping*), *spelling normalization*, *case folding*, *tokenizing*, *stopword*, *stemming* dan klasifikasi opini menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Hasil dari penelitian ini didapatkan bahwa pasangan Capres 01 mendapatkan jumlah 14,2% untuk sentimen positif, 68,15% sentimen negatif dan 19,5% sentimen netral. Sedangkan pasangan capres 02 mendapatkan total 24,7% sentimen positif, 64,5% sentimen negatif, dan 10,65% sentimen netral. Algoritma *Naïve Bayes Classifier* merupakan algoritma yang tepat untuk kasus analisis sentimen masyarakat media sosial *Twitter* dengan akurasi 94% dan 99% sedangkan pada metode *Support Vector Machine* hanya memperoleh akurasi 72% dan 65%.

## KATA PENGANTAR

Segala puji bagi Allah SWT. Tuhan semesta alam yang terus menerus melimpahkan rahmat serta memberikan kekuatan dan keimanan kepada kita semua selaku makhluk ciptaan-Nya. Shalawat serta salam semoga tercurah limpahkan kepada Nabi akhir zaman baginda Nabi Muhammad SAW, kepada keluarganya, sahabatnya, sampai kepada kita selaku umatnya.

Atas izin Allah SWT penulis dapat menyelesaikan Hasil Proposal tentang “Analisis Sentimen Masyarakat Terhadap Peserta Pemilu Pada Media Sosial *Twitter* Dengan Perbandingan Metode *Naive Bayes Classifier* Dan *Support Vector Machine*”. Penulis dengan senang hati ingin mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Dr. Prihastuti Harsani, M.Si. sebagai Pembimbing I yang telah memberikan bimbingan serta motivasi kepada penulis dalam penyusunan hasil proposal ini.
2. Irma Anggraeni, M.Kom. sebagai Pembimbing II yang telah memberikan bimbingan serta motivasi kepada penulis dalam penyusunan hasil proposal ini.
3. Prihastuti Harsani, M.Si, selaku Kepala Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam.
4. Kedua Orang Tua yang telah memberikan dukungan, motivasi dan doa yang terus menerus untuk keberhasilan penyusunan laporan ini.
5. Semua Team Analisis dan rekan kelas C dan D Ilmu Komputer angkatan 2015 yang selalu memberikan *support* dan bantuan dalam pengerjaan laporan ini.
6. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung yang tidak dapat disebutkan satu persatu terimakasih atas segalanya.

Disadari Proposal ini masih kurang sempurna. Oleh karena itu, kritik dan saran dari semua pihak sangat diharapkan guna penyempurnaan Proposal ini. Semoga Proposal ini bermanfaat bagi pembacanya.

Bogor, 05 Agustus 2019

Fany Syafira Nurul Kilah

0651 15 125



## DAFTAR ISI

<b>HALAMAN PENGESAHAN.....</b>	<b>i</b>
<b>PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI.....</b>	<b>ii</b>
<b>RIWAYAT HIDUP.....</b>	<b>iv</b>
<b>RINGKASAN.....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR TABEL.....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>x</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Ruang Lingkup.....	2
1.3. Tujuan.....	2
1.4. Manfaat.....	2
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>3</b>
2.1. Twitter.....	3
2.2. Pemilu.....	3
2.3. Text Mining.....	3
2.3.1. Text Preprocessing.....	4
2.3.2. Feature Selection.....	4
2.4. Sentiment Analysis.....	4
2.5. Naive Bayes Classifier.....	4
2.6. Support Vector Machine.....	5
2.7. Software R.....	6
2.8. Algoritma Enhanced Confix Stripping (ECS) Stemmer.....	6
2.9. Penelitian Terdahulu.....	6
<b>BAB III METODE PENELITIAN.....</b>	<b>9</b>

3.1. Requirement Analysis (Analisis) .....	9
3.2. System Design (Desain) .....	9
3.3. Implementation (Implementasi) .....	9
3.4. Integration & Testing (Uji Coba) .....	9
3.5. Waktu dan Tempat Penelitian.....	10
3.6. Alat dan Bahan.....	10
<b>BAB IV PERANCANGAN DAN IMPLEMENTASI.....</b>	<b>11</b>
4.1. Tahap Requirement Analysis (Analisis) .....	11
4.1.1. Tahap Analisis Data.....	11
4.2. System Design (Desain) .....	12
4.2.1. Perancangan Sistem) .....	12
4.3. Implementation (Implementasi) ) .....	21
4.3.1. Proses Scraping.....	22
4.3.2. Proses Text Mining.....	22
4.3.3. Proses Klasifikasi.....	22
4.4. Integration & Testing (Uji Coba).....	23
<b>BAB V HASIL DANPEMBAHASAN.....</b>	<b>24</b>
5.1. Hasil.....	24
5.2. Pembahasan.....	29
5.2.1. Hasil Analisis Sentimen.....	29
5.2.2. Hasil Akurasi Confusion Martix.....	32
<b>BAB VI KESIMPULAN DAN SARAN.....</b>	<b>33</b>
6.1. Kesimpulan.....	33
6.2. Saran.....	33
<b>DAFTAR PUSTAKA.....</b>	<b>34</b>
<b>LAMPIRAN.....</b>	<b>xii</b>
Lampiran 1. Kartu Bimbingan.....	xiii
Lampiran 2. Kartu Hadir Seminar.....	xiv
Lampiran 3. Surat Keterangan Maju Sidang Skripsi.....	xv

## DAFTAR TABEL

Tabel 1. Tabel Perbandingan Penelitian.....	8
Tabel 2. Jadwal Penelitian.....	10
Tabel 3. Tabel Alat dan Bahan.....	10
Tabel 4. Data Keyword Scraping.....	11
Tabel 5. Tabel Stopword.....	11
Tabel 6. Tabel Kata Dasar.....	12
Tabel 7. Tabel Data Awal.....	19
Tabel 8. Tabel Normalisasi Data.....	20
Tabel 9. Tabel Heusian.....	20
Tabel 10. Tabel Pelabelan Sentimen.....	21
Tabel 11. Tabel Hasil Analisis Sentimen.....	31
Tabel 12. Tabel Akurasi.....	32

## DAFTAR GAMBAR

Gambar 1. Metode Waterfall.....	9
Gambar 2. Flowchart Program.....	12
Gambar 3. Spelling Normalization.....	13
Gambar 4. Case Folding.....	14
Gambar 5. Tokenizing.....	14
Gambar 6. Filtering.....	15
Gambar 7. Flowchart Text Mining.....	15
Gambar 8. Proses Scraping.....	22
Gambar 9. Proses Pengolahan Data dan Kata.....	22
Gambar 10. Proses Klasifikasi dengan 2 Metode.....	23
Gambar 11. Proses Instalasi Package.....	24
Gambar 12. Proses Pemanggilan Library.....	25
Gambar 13. Proses Scraping Data.....	25
Gambar 14. Proses Text Mining.....	26
Gambar 15. Proses Pembacaan Corpus dan Pelabelan Score .....	27
Gambar 16. Proses Pembuatan Grafik.....	27
Gambar 17. Proses Klasifikasi dengan Metode NBC .....	28
Gambar 18. Proses Klasifikasi dengan Metode SVM.....	28
Gambar 19. Grafik Analisis Sentimen berbahasa indonesia Capres No Urut 1.....	29
Gambar 20. Grafik Analisis Sentimen berbahasa Indonesia Capres No Urut 2.....	30
Gambar 21. Grafik Analisis Sentimen berbahasa Inggris Capres No Urut 1.....	31
Gambar 22. Grafik Analisis Sentimen berbahasa Inggris Capres No Urut 2.....	32

## DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan.....	xiii
Lampiran 2. Kartu Hadir Seminar.....	xiv
Lampiran 3. Surat Keterangan Maju Sidang Skripsi.....	xv

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Situs *microblogging* telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet. Jutaan pesan yang muncul setiap hari di situs *web* populer yang menyediakan layanan *microblogging* seperti *Twitter*, *Tumblr*, dan *Facebook*. Hal tersebut menyebabkan semakin banyak pengguna yang melakukan posting tentang suatu produk dan layanan yang mereka gunakan, atau mengekspresikan pandangan mereka tentang politik dan agama. *Twitter* sebagai salah satu situs *microblogging* dengan pengguna lebih dari 500 juta dan 400 juta *tweet* perhari memungkinkan pengguna untuk berbagi pesan menggunakan teks pendek disebut *tweet*. Data tersebut dapat digunakan secara efisien untuk pemasaran atau studi sosial. *Sentiment analysis* atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral.

Informasi yang terkandung dalam *Twitter* ini sangat berharga sebagai alat penentu kebijakan. Hal ini bisa dilakukan dengan menggunakan analisis sentimen karena dengan mengetahui apa yang menyebabkan sentimen berubah, pihak yang bersangkutan bisa mengambil keputusan dengan lebih baik. Sedangkan ketika mengetahui suatu topik atau kejadian menyebabkan sentimen turun, maka pihak yang bersangkutan akan menghindari kejadian serupa untuk meningkatkan sentimen. Kebutuhan-kebutuhan tersebut biasanya muncul ketika suatu pihak ingin mendapatkan sentimen publik yang baik atau melakukan pencitraan

Berdasarkan permasalahan tersebut dibutuhkan sebuah program yang dapat menganalisis sentimen masyarakat pada *Twitter* terhadap suatu topik mengenai isu politik peserta pemilu yang akan ditampilkan dalam bentuk grafik agar lebih mudah dibaca dan dipahami.

Penelitian berikutnya dilakukan oleh Faishol Nurhuda, Sari Widya Sihwi dan Afrizal Doewes pada tahun 2013. Dalam penelitian ini akan dilakukan analisis sentimen masyarakat terhadap calon presiden dan wakil presiden Indonesia 2014 yang diungkapkan melalui jejaring sosial *Twitter*. Ada beberapa tahap untuk melakukan analisis sentimen, diantaranya adalah tahap pengumpulan data, *praprocessing* data dan klasifikasi opini menggunakan metode *Naïve Bayes Classifier*. Hasil dari penelitian ini didapatkan bahwa Prabowo Subianto – Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 26,4% sentimen negatif dan 25,9% sentimen netral. Sedangkan pasangan Joko Widodo – Jusuf Kalla mendapatkan total 37,6% sentimen positif, 34,4% sentimen negatif, dan 27,9 sentimen netral.

Penelitian berikutnya dilakukan oleh Taufik dan S.A. Pamungkas pada tahun 2018. Penelitian ini telah berhasil mengimplementasikan Algoritma *SVM* untuk analisa sentimen pada data *tweet* tentang tokoh publik dengan keyword “Ahok” dan “@teman\_ahok”. Hal ini diperlihatkan dengan hasil akurasi menggunakan data *tweet* sebanyak 630 data, dimana 420 data adalah data *training* dan 210 adalah data *testing*.

Hasil menunjukkan bahwa *Kernel Linear* memiliki tingkat presisi paling baik sekitar 80% dilanjutkan dengan *Kernel Sigmoid* sekitar 75%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%. *Kernel sigmoid* memiliki tingkat *recall* paling baik sekitar 85% dilanjutkan dengan *Kernel Linear* sekitar 76%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%. *Kernel Sigmoid* memiliki tingkat akurasi paling baik sekitar 81% dilanjutkan dengan *Kernel Linear* sekitar 77%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%.

Penelitian selanjutnya dilakukan oleh Adhi Viky Sudiantoro dan Eri Zuliarso pada tahun 2018. Dalam penelitian ini menggunakan teks Bahasa Indonesia yang terdapat di sosial media twitter berupa *tweet*. Data yang digunakan terdiri 300 data *tweet* yang dibagi menjadi 2 yaitu untuk data latih sebanyak 200 data dan 100 data untuk data uji. Pengklasifikasian data *tweet* menggunakan *text mining* dengan *Naïve Bayes Classifier*. Sebelum klasifikasi, dilakukan beberapa tahap pemrosesan teks seperti *case folding*, *normalisasi*, *tokenisasi* dan *stopwords removal*. Hasil dari 100 data uji yang klasifikasi menghasilkan 32 data bersentimen positif dan sebanyak 68 data sentimen negatif. Dapat diartikan bahwa 100 data uji yang diklasifikasi masuk dalam kategori bersentimen negatif dikarenakan data positif lebih kecil daripada data yang bersentimen negatif. Adapun nilai akurasi algoritma *Naïve Bayes Classifier* memberikan nilai akurasi sebesar 84%.

## 1.2. Ruang Lingkup

Ruang lingkup penelitian ini berdasarkan data *Twitter* yang diambil pada bulan April 2019 sebanyak 600 data berbahasa Indonesia dan berbahasa Inggris mengenai sentimen masyarakat terhadap peserta pemilu (calon presiden 2019) yang dikelompokkan menjadi tiga sentimen yaitu positif, negatif, dan netral dengan menggunakan metode *Naive Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)* yang kemudian akan dibandingkan keakuratannya. Hasil dari data tersebut akan ditampilkan dalam bentuk grafik dan pemrograman yang digunakan dalam penelitian ini yaitu R.

## 1.3. Tujuan

Penelitian ini diharapkan dapat memberi tambahan pengetahuan tentang implementasi algoritma *Naive Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)* yang dapat memberikan gambaran mengenai bagaimana analisis sentimen pada jejaring sosial *Twitter* untuk mengklasifikasi sentimen untuk berbagai kepentingan serta pengoptimalan informasi jejaring sosial untuk kepentingan publik. Selain itu juga bermanfaat untuk mencari informasi tentang sentimen masyarakat mengenai calon presiden 2019 apakah mempunyai respon positif netral atau negatif dari masyarakat dan juga bermanfaat bagi tokoh publik untuk melakukan pengukuran respon masyarakat terhadap dirinya.

#### **1.4. Manfaat**

Manfaat penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan sentimen pada *Twitter* dalam jumlah besar secara otomatis.
2. Mencari informasi dan menentukan apakah calon presiden 2019 tersebut dilihat positif, negatif atau netral di twitter.
3. Mengetahui metode paling akurat untuk studi kasus analisis sentimen pada *twitter*.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. *Twitter*

*Twitter* adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh *Twitter Inc.* yang menawarkan jaringan sosial berupa *mikroblog* sehingga memungkinkan penggunaannya untuk mengirim dan membaca pesan *tweets*. *Microblog* adalah salah satu jenis alat komunikasi *online* dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu. *Tweets* adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. *Tweets* bisa dilihat secara publik, namun pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat melihat *tweets* pengguna lain yang dikenal dengan sebutan pengikut (*follower*). Tidak seperti *Facebook*, *LinkedIn*, dan *MySpace*. *Twitter* merupakan sebuah jejaring sosial yang dapat digambarkan sebagai sebuah *graph* berarah yang berarti bahwa pengguna dapat mengikuti pengguna lain, namun pengguna kedua tidak diperlukan untuk mengikutinya kembali. Kebanyakan akun berstatus publik dan dapat diikuti tanpa memerlukan persetujuan pemilik. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda # (*hashtag*). Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda @. (Boy, 2014).

#### 2.2. Pemilu

Pemilu di Indonesia pada mulanya memiliki tujuan untuk memilih anggota DPR, baik di tingkat pusat maupun di daerah (Provinsi, Kabupaten/Kota). Setelah amandemen ke-4 UUD Negara Republik Indonesia Tahun 1945, pemilu juga bertujuan untuk memilih presiden dan wakil presiden yang selanjutnya dikenal sebagai Pilpres. Pilpres yang pada awalnya dipilih oleh MPR, kemudian dirubah menjadi sistem pemilihan langsung oleh rakyat. Pilpres yang dilaksanakan secara langsung pertama kali diadakan di Tahun 2004. Setelah adanya UU Nomor 22 Tahun 2007, pilkada (pemilihan kepala daerah) juga masuk kedalam rangkaian pesta demokrasi di Indonesia. Pemilihan umum berlangsung setiap 5 tahun sekali. Di tahun 2009, pemilu dilaksanakan pada tanggal 9 april 2009, ada 560 anggota DPR serta 132 Anggota DPD. Sekitar 38 partai mengikuti pesta demokrasi ini. (Khoiril, 2018).

#### 2.3. *Text Mining*

*Text mining* (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007).

Tahap-tahap *text mining* menurut Feldman & Sanger 2007 secara umum adalah *text preprocessing* dan *feature selection*. Dimana penjelasan dari tahap-tahap tersebut adalah sebagai berikut :

### **2.3.1. Text Preprocessing**

Tahap *text preprocessing* adalah tahap awal dari *text mining*. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* sistem text mining. Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter - delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut (T Zhang, 2005).

### **2.3.2. Feature Selection**

Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan. Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang berlebihan. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stopword* (*stoplist*). Jika termasuk di dalam stoplist maka kata-kata tersebut akan dihapus dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang mencirikan isi dari suatu dokumen atau *keywords*. Setelah melalui proses *stopword removal* tindakan selanjutnya adalah yaitu proses *stemming*. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). Tujuan dari proses stemming adalah menghilangkan imbuhan-imbuhan baik itu berupa *prefiks*, *sufiks*, maupun *konfiks* yang ada pada setiap kata. Jika imbuhan tersebut tidak dihilangkan maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang melekatinya sehingga hal tersebut akan menambah beban *database*. Hal ini sangat berbeda jika menghilangkan imbuhan-imbuhan yang melekat dari setiap kata dasar, maka satu kata dasar akan disimpan sekali walaupun mungkin kata dasar tersebut pada sumber data sudah berubah dari bentuk aslinya dan mendapatkan berbagai macam imbuhan. Karena bahasa Indonesia mempunyai aturan morfologi maka proses stemming harus berdasarkan aturan morfologi bahasa Indonesia. (Tala, 2003).

## **2.4. Sentiment Analysis**

*Sentiment analysis* 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.

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudia menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Bing, 2012).

## 2.5. Naive Bayes Classifier

Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger 2007). Dalam penelitian ini yang menjadi data uji adalah dokumen *tweets*. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Proses pembelajaran *Naive Bayes Classifier* pada perhitungan setiap kelas yang ada. Rumusnya adalah sebagai berikut :

$$P(V_j) = \frac{N_c}{N} \dots \dots \dots (1)$$

Keterangan :

- $P(V_j)$  : Probabilitas kelas yang dimiliki
- $N_c$  : Jumlah dari kelas yang ada
- $N$  : Jumlah dari data *training*

Tahap berikutnya melakukan perhitungan pada kata yang terdapat data test dan rumusnya adalah sebagai berikut :

$$P(x_i|V_j) = \frac{nk+1}{n+|Vocabulary|} \dots \dots \dots (2)$$

Keterangan :

- $P(x_i|V_j)$  : Probabilitas  $x_i$  pada kelas  $V_j$ ,  $x_i$  adalah kata dalam data test dan  $V_j$  kelas yang ada seperti positif dan negatif
- $nk$  : Jumlah kata yang terdapat pada setiap kelas
- $n$  : Jumlah data training pada kata
- $Vocabulary$  : Jumlah semua kata dari semua kategori

Proses klasifikasi *Naive Bayes Classifier* secara umum menentukan kelas dari data test dengan menggunakan rumus sebagai berikut :

$$P(x_i|V_j) = \underset{V_j \in V}{\arg \max} \prod_{i=1}^n P(x_i|V_j) \dots \dots \dots (3)$$

## 2.6. Support Vector Machine

*Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training. Dengan melakukan training menggunakan data inputan dalam bentuk numerik dan pembobotan dengan akan didapatkan sebuah pola yang nantinya akan digunakan dalam proses pelabelan. Nilai atau pola yang dihasilkan dari *Metode Support Vector Machine* sebenarnya adalah sebuah garis pemisah yang disebut dengan *Hyperplane*, dimana pada kasus ini garis tersebut berperan dalam memisahkan tweet dengan sentimen positif dengan tweet yang memiliki sentimen negatif . (Novantirani,2015)

## 2.7. Software R

Versi pertama R diluncurkan pada tahun 1992 oleh Ross Ihaka dan Robert Gentleman (1996) (singkatan R berasal dari kedua nama tersebut) yang keduanya dari *The University of Auckland*. Diawal pengembangannya, proyek R dibuat dengan bahasa LISP serta hanya diimplementasikan di *macintosh* dengan bahasa semantik *Scheme*. Pada saat ini *source code* R yang dibuat telah bersifat *multiplatform*, sehingga dapat dikompilasi dan dijalankan diberbagai sistem operasi berbasis *\*NIX* dan *Windows*. R dikembangkan secara intensif oleh *R-core Team* yang anggotanya terdiri dari 17 orang ahli statistika dan melibatkan banyak kontributor lain dari berbagai institusi diseluruh dunia. Seperti halnya *linux*, *source code* R tersedia untuk dibaca, dianalisis, identifikasi *bug* dan modifikasi. Bahasa R berbasis bahasa S yang dibangun di *Bell laboratories* di tahun 80-an. Dengan kata lain R adalah implementasi bahasa S yang sudah dibangun oleh Rick Becker, John Chambers dan Allan Wilks, yang juga membentuk dasar dari sistem *S-Plus*. Dengan demikian sintak R hampir identik dengan S dengan perbedaan yang tidak terlalu banyak. Sebagai contoh mulai R versi 1.7.0 *assign* variabel diperbolehkan menggunakan “=” yang mana di S *assign* variabel menggunakan “<-”. (Rosiiter,2009)

## 2.8. Algoritma Enhanced Confix Stripping (ECS) Stemmer

*Stemming* merupakan bagian dari proses *Information Retrieval* (IR), yang mengubah beberapa kata ke bentuk kata dasarnya sebelum dilakukan pengindeksan. Proses *stemming* bekerja tergantung pada bahasa yang diteliti. Khusus untuk topik berbahasa Indonesia, proses *stemming* awalnya bekerja berdasarkan struktural morfologi kalimat bahasa Indonesia, yang terdiri dari prefiks (awalan), sufiks (akhiran), infiks (sisipan), dan konfiks (awalan+akhiran). Algoritma terbaru yang diteliti oleh Mahendra, selanjutnya dinamakan algoritma *Enhanced Confix Stripping (ECS) Stemmer*. Algoritma ini merupakan pengembangan dari algoritma *Confix Stripping (CS) Stemmer*, dan berhasil mereduksi jumlah term pada algoritma *Confix Stripping Stemmer* hingga 32.66%, sedangkan pada awalnya *Confix Stripping Stemmer* hanya mampu mereduksi 30.95% term. (Ade, Maya & Romi, 2014)

## 2.9. Penelitian Terdahulu

- 1 Nama : Faishol Nurhuda, Sari Widya Sihwi, Afrizal Doewes  
Judul : Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier  
ISBN : 2301-7201  
Isi : Penelitian berikutnya dilakukan oleh Faishol Nurhuda, Sari Widya Sihwi dan Afrizal Doewes pada tahun 2013. Dalam penelitan ini akan dilakukan analisis sentimen masyarakat terhadap calon presiden dan wakil presiden Indonesia 2014 yang diungkapkan melalui jejaring sosial Twitter. Ada beberapa tahap untuk melakukan analisis sentimen, diantaranya adalah tahap pengumpulan data, *praprocesing* data dan klasifikasi opini menggunakan metode *Naive Bayes Classifier*. Hasil dari penelitian ini didapatkan bahwa Prabowo Subianto – Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 26,4% sentimen negatif dan 25,9% sentimen netral. Sedangkan

pasangan Joko Widodo – Jusuf Kalla mendapatkan total 37,6% sentimen positif, 34,4% sentimen negatif, dan 27,9 sentimen netral.

- 2 Nama : Taufik, S.A. Pamungkas  
Judul : Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (Svm)  
ISSN : 1978 – 8568  
Isi : Penelitian ini telah berhasil mengimplementasikan Algoritma SVM untuk analisa sentimen pada data tweet tentang tokoh publik dengan keyword “Ahok” dan “@teman\_ahok”. Hal ini diperlihatkan dengan hasil akurasi menggunakan data *tweet* sebanyak 630 data, dimana 420 data adalah data *training* dan 210 adalah data *testing*. Hasil menunjukkan bahwa *Kernel Linear* memiliki tingkat presisi paling baik sekitar 80% dilanjutkan dengan *Kernel Sigmoid* sekitar 75%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%. *Kernel sigmoid* memiliki tingkat *recall* paling baik sekitar 85% dilanjutkan dengan *Kernel Linear* sekitar 76%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%. *Kernel Sigmoid* memiliki tingkat akurasi paling baik sekitar 81% dilanjutkan dengan *Kernel Linear* sekitar 77%. Untuk *Kernel Gaussian* dan *Polinom* nilai akurasinya sekitar 50%.
  
- 3 Nama : Adhi Viky Sudiantoro dan Eri Zuliarso  
Judul : Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier  
ISSN : 978-602-8557-20-7  
Isi : Dalam penelitian ini menggunakan teks Bahasa Indonesia yang terdapat di sosial media twitter berupa *tweet*. Data yang digunakan terdiri 300 data *tweet* yang dibagi menjadi 2 yaitu untuk data latih sebanyak 200 data dan 100 data untuk data uji. Pengklasifikasian data *tweet* menggunakan *text mining* dengan *Naïve Bayes Classifier*. Sebelum klasifikasi, dilakukan beberapa tahap pemrosesan teks seperti *case folding*, *normalisasi*, *tokenisasi* dan *stopwords removal*. Hasil dari 100 data uji yang klasifikasi menghasilkan 32 data bersentimen positif dan sebanyak 68 data sentimen negative. Dapat diartikan bahwa 100 data uji yang diklasifikasi masuk dalam kategori bersentimen negatif dikarenakan data positif lebih kecil daripada data yang bersentimen negatif. Adapun nilai akurasi algoritma Naïve Bayes Classifier memberikan nilai akurasi sebesar 84%.
  
- 4 Nama : Ahmad Fathan Hidayatullah, Azhari SN  
Judul : Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter

ISSN : 1979-2328

Isi : Penelitian ini telah berhasil membangun model untuk melakukan klasifikasi *tweet* berdasarkan sentimen dan kategori dengan *Naive Bayes Classifier*. Hasil akurasi pengujian klasifikasi dengan fitur *term frequency* diperoleh sebesar 79,91% sedangkan fitur *TF-IDF* didapatkan akurasi sebesar 79,68%. Klasifikasi menggunakan tools *RapidMiner* dengan *Naive Bayes* dan fitur *term frequency* diperoleh sebesar 73,81% sedangkan dengan fitur *TF-IDF* diperoleh sebesar 71.11%. Klasifikasi dengan *Support Vector Machine* menghasilkan akurasi 83,14% untuk fitur *term frequency* dan 82,69% untuk fitur *TF-IDF*. Hasil pengujian pada aplikasi yang dibangun dan pada tools *RapidMiner* memperlihatkan bahwa akurasi dengan fitur *term frequency* memberikan hasil akurasi yang lebih baik daripada akurasi dengan fitur *TF-IDF*. Metode *Support Vector Machine* menghasilkan akurasi performansi yang lebih baik daripada metode *Naive Bayes*. Penggunaan metode *Support Vector Machine* dan *Naive Bayes* sama-sama memiliki hasil akurasi yang cukup baik untuk klasifikasi *tweet*.

5 Nama : Agung Pramono , Rini Indriati, Arie Nugroho

Judul : Sentiment Analysis Tokoh Politik Pada Twitter

ISSN : 2549-7952

Isi : Proses pengklasifikasian metode *Naive Bayes Classifier* ada beberapa tahap. Tahap pertama Proses Pembelajaran *Naive Bayes Classifier* tahap kedua Proses Klasifikasi *Naive Bayes Classifier*. Hasil proses klasifikasi berdasarkan nilai *Vmap* diambil nilai terbesar yang menjadi kategori dari *tweet* kelas positif, negatif dan netral. Pada proses klasifikasi nilai akurasi tergantung banyaknya data *training*. Pada data *training* positif 100, negatif 100, netral 100 dan data *test* 100, nilai akurasi 61% dan error 39% . Pada data *training* positif 700, negatif 700 dan netral 700 dengan menggunakan data *test* 100 akurasinya 91% dan *error* 9%. Rekomendasi penelitian selanjutnya menggunakan bahasa daerah atau bahasa asing, mengklasifikasi tentang produk atau hal lain, mengklasifikasi *tweet* yang terdapat singkatan.

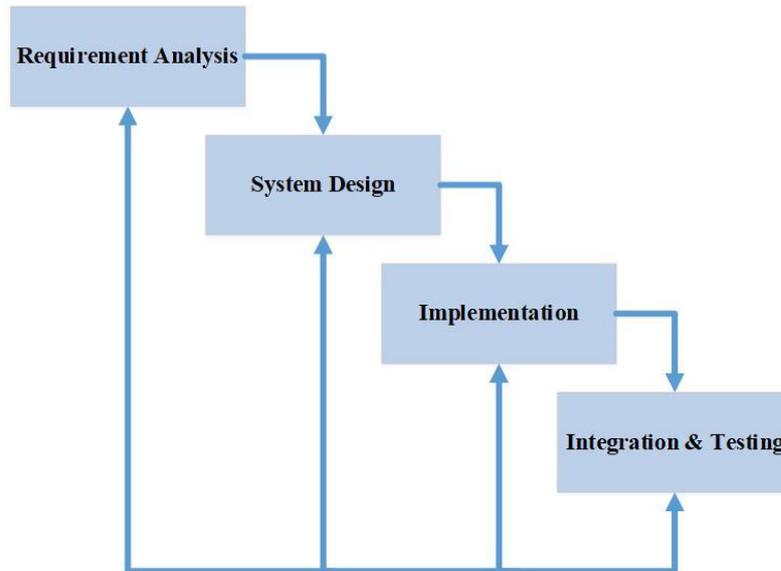
**Tabel 1. Tabel Perbandingan Penelitian**

No	Nama (Tahun)	Metode		Pemrograman		Aplikasi	
		NBC	SVM	PHP	R	Dekstop	Web
1	Faishol Nurhuda, Sari Widya Sihwi, Afrizal Doewes (2014)	✓		✓			✓
2	Taufik, S.A. Pamungkas (2018)		✓	✓			✓
3	Adhi Viky Sudiantoro dan Eri Zuliarso (2018)	✓		✓			✓
4	Ahmad Fathan Hidayatullah, Azhari SN (2014)	✓	✓	✓			✓
5	Agung Pramono, Rini Indriati, Arie Nugroho (2017)	✓		✓			✓
6	Fany Syafira Nurul Kilah (2019)	✓	✓		✓	✓	

## BAB III

### METODE PENELITIAN

Dalam pengembangannya metode *waterfall* memiliki beberapa tahapan yang berurut yaitu *requirement analysis* (analisis kebutuhan), *design system* (desain sistem), *coding* (pengkodean) *testing* (pengujian) dan penerapan program. Tahapan tahapan dari metode *waterfall* adalah sebagai berikut :



Gambar 1. Metode *Waterfall*

#### **3.1. Requirement Analysis (Analisis)**

Tahap ini pengembang sistem diperlukan komunikasi yang bertujuan untuk memahami perangkat lunak yang diharapkan oleh pengguna dan batasan perangkat lunak tersebut. Informasi ini biasanya dapat diperoleh melalui wawancara, diskusi atau survei langsung. Informasi dianalisis untuk mendapatkan data yang dibutuhkan oleh pengguna.

#### **3.2. System Design (Desain)**

Tahapan ini bertujuan untuk memberikan gambaran mengenai rancangan program atau desain yang akan dikembangkan.

### 3.3. Implementation (Implementasi)

Pada tahap ini, pertama kali dikembangkan di program kecil yang disebut unit, yang terintegrasi dalam tahap selanjutnya. Setiap unit dikembangkan dan diuji untuk fungsionalitas yang disebut sebagai *unit testing*.

### 3.4. Integration & Testing (Uji Coba)

Seluruh *unit* yang dikembangkan dalam tahap implementasi diintegrasikan ke dalam sistem setelah pengujian yang dilakukan masing-masing program.

### 3.5. Waktu dan Tempat Penelitian

Penelitian ini akan dilaksanakan pada bulan April sampai Juni 2019, jadwal penelitian akan ditampilkan lebih rinci pada tabel 2 dan tempat penelitian ini akan dilaksanakan di Laboratorium Multimedia Program Studi Ilmu Komputer Universitas Pakuan Bogor.

**Tabel 2. Jadwal Penelitian**

No	Kegiatan	Bulan											
		April				Mei				Juni			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Requirement Analisis	■	■	■	■								
2	System Design					■	■	■	■				
3	Implementation									■	■		
4	Integration & Testing											■	■
5	Pembuatan Laporan	■	■	■	■	■	■	■	■	■	■	■	■

### 3.6. Alat dan Bahan

Alat yang digunakan dalam penelitian ini terdiri dari perangkat keras (Hardware) dan perangkat lunak (Software).

**Tabel 3. Tabel Alat dan Bahan**

<b>Alat</b>		<b>Bahan</b>
<b>Hardware</b>	<b>Software</b>	
Aspire E1-410	Xampp Control Panel v3.2.2	Buku dan Jurnal sebagai referensi
ROM 500 GB HDD	Visio Profesional 2013	Data Twitter pada bulan april mengenai Capres 2019
RAM 4 GB DDR3 Memory	R	
Processor Intel (R) Celeron (R) quad core processor N2920 (2M Cache, 2.0 GHz)	Mozila Firefox	

## BAB IV

### PERANCANGAN DAN IMPLEMENTASI

#### 4.1. Tahap *Requirement Analysis* (Analisis)

Tahap ini bertujuan untuk memahami perangkat lunak yang diharapkan oleh pengguna dan batasan perangkat lunak tersebut. Informasi ini didapat dengan memanfaatkan *API Twitter*. Informasi dianalisis untuk mendapatkan data yang dibutuhkan oleh pengguna

##### 4.1.1. Tahap Analisis Data

Tahap analisis data merupakan tahapan untuk memperoleh data, mempersiapkan korpus, dan perancangan database.

###### 4.1.1.1. Data *Twitter*

Data *Tweet* dalam penelitian ini diperoleh dengan *scraping* data yang disediakan oleh *Twitter*. Dengan memanfaatkan *software R* tersebut akan didapat sebuah file dengan format *XML (file excel)*. Pada saat pengumpulan data akan dimasukan *keyword-keyword* yang berhubungan dengan calon presiden. Berikut ini adalah daftar kata-kata yang digunakan sebagai kata kunci pencarian data.

**Tabel 4. Data Keyword Scraping**

Nomer Urut 1	Nomer Urut 2
#jokowi	#prabowo
#jokowidodo	#prabowosubianto
#jokowimaruf	#prabowosandi

###### 4.1.1.2. Data *Stopwords*

Data *stopword* didapat dari dari *tweet* yang digunakan dalam penelitian. Rancangan tabel kata *stopword* dapat dilihat pada tabel dibawah ini :

**Tabel 5. Tabel Stopword**

Jenis	Stopword
Kata	Ada, adalah, sebuah, ketika, mengapa, bagaimana, apa, sedang, dan, juga, jika, kapan, siapa, dimana, kenapa, dan, dll
Karakter	. , ? / > < “ ‘ : :   \ + = _ - ) ( * & ^ % \$ # @ ! ~ ` 1 2 3 4 5 6 7 8 9 0 [ ] { }
Bahasa Tweet	Link, RT, nama akun, nama hastag, & amp

#### 4.1.1.3. Data Kata Dasar

Data kata dasar didapat dari kamus bahasa Indonesia online dimana datanya berjumlah 28533 data. Data kata dasar disimpan di dalam database. Rancangan tabel kata dasar dapat dilihat pada tabel 6 :

**Tabel 2. Tabel Kata Dasar**

Id	KATA DASAR
1	Ada
2	Adalah
3	Dll

## 4.2. System Design (Desain)

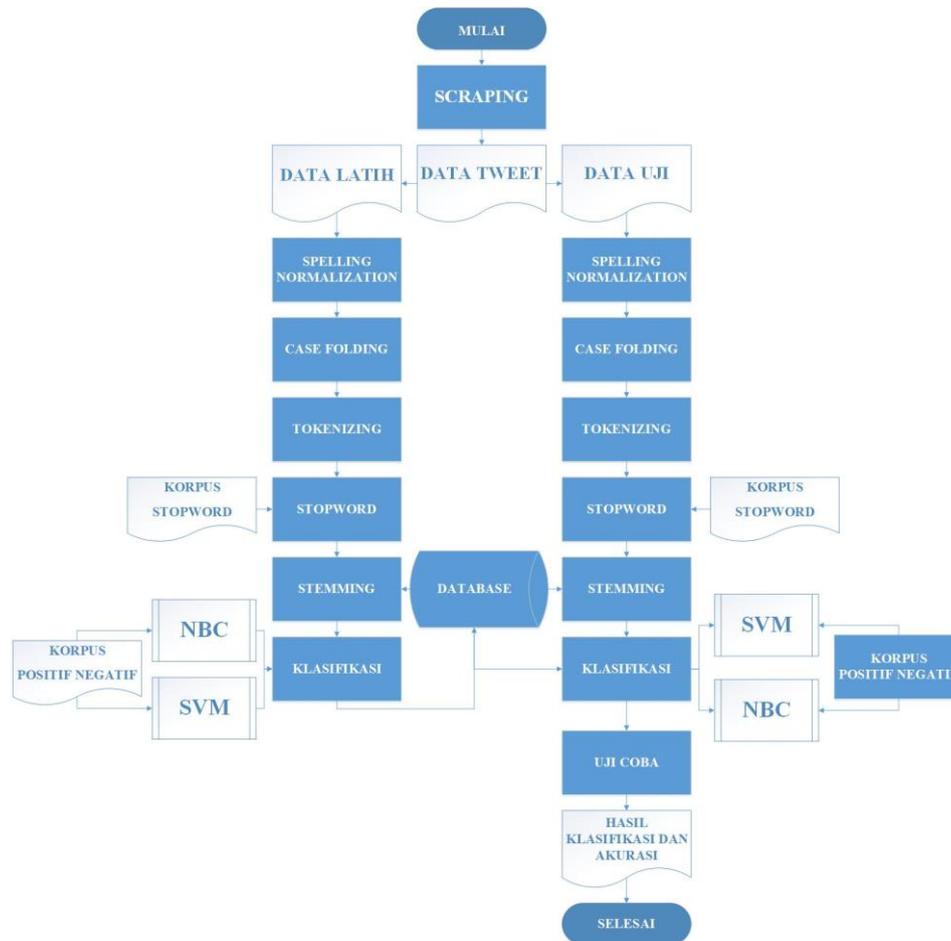
Tahapan ini bertujuan untuk memberikan gambaran mengenai rancangan sistem atau desain yang akan dikembangkan.

### 4.2.1. Perancangan Sistem

Perancangan sistem dibuat dan digunakan untuk mempermudah dalam pembuatan sistem sebab dalam perancangan sistem secara detail ini berisi *flowchart program*

#### 4.2.1.1. Flowchart Program

Dalam perencanaan sistem, diagram alur atau *Flowchart Program* digunakan untuk menggambarkan proses kerja. Pada gambar 2 ini adalah *Flowchart Program* dari analisis sentimen peserta pemilu.



Gambar 2. Flowchart Program

#### 4.2.1.2. Text Preprocessing

Sebelum melakukan klasifikasi pada dokumen teks, perlu dilakukan *preprocessing*. Data ulasan yang diperoleh belum sepenuhnya siap digunakan untuk proses klasifikasi secara langsung karena data masih tidak terstruktur dengan baik dan terdapat banyak *noise*. Data masih memuat angka, tanda baca serta kata-kata lain yang kurang bermakna untuk dijadikan fitur. Maka dari itu, perlu dilakukan *preprocessing* yang bertujuan untuk menyeragamkan bentuk kata, menghilangkan karakter-karakter selain huruf, dan mengurangi volume kosakata sehingga data akan lebih terstruktur. Pada tahap *preprocessing*, akan dilakukan pembersihan data menggunakan metode *text mining*. Beberapa tahap yang akan dilakukan diantaranya adalah *spelling normalization*, *case folding*, *tokenizing*, dan *filtering* yang akan dijelaskan pada sub bab berikut :

#### 4.2.1.3. Spelling Normalization

*Spelling* adalah tahap awal yang perlu dilakukan untuk mendapatkan kualitas dokumen yang baik. *Spelling* melakukan perbaikan kata-kata yang salah eja atau disingkat dengan bentuk tertentu. Pada penelitian ini proses *spelling* dilakukan dengan

bantuan Microsoft Excel dan software R. Contoh penerapan fitur spelling dapat dilihat pada Gambar .



Gambar 3. *Spelling Normalization*

#### 4.2.1.4. *Case Folding*

*Case folding* adalah proses penyeragaman bentuk huruf dimana dalam proses ini hanya menerima huruf latin antara “a” sampai “z”. Karakter lain selain huruf dianggap sebagai delimiter sehingga karakter tersebut akan dihapus dari dokumen. Kemudian penyeragaman dilakukan dengan mengubah isi dokumen menjadi huruf kecil secara keseluruhan (dari “a” sampai dengan “z”). Hal ini bertujuan agar kata yang ditulis dengan huruf awal kapital dan huruf non kapital tidak terdeteksi memiliki arti yang berbeda. Contoh penggunaan *case folding* dapat dilihat contoh pada gambar 4 berikut :

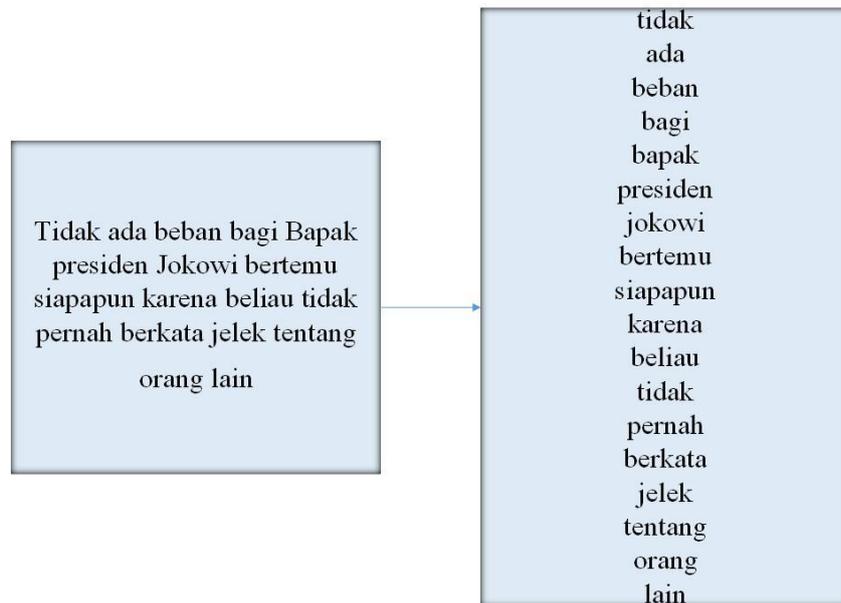


Gambar 4. *Case Folding*

#### 4.2.1.5. *Tokenizing*

*Tokenizing* atau tokenisasi adalah proses memisahkan kata per kata pada sebuah dokumen menjadi kata – kata yang saling independen. *Tokenizing* dilakukan untuk mendapatkan token atau potongan kata yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Tokenisasi dapat memudahkan proses perhitungan keberadaan kata tersebut dalam dokumen ataupun

untuk menghitung frekuensi kemunculan kata tersebut dalam corpus. Contoh proses *tokenizing* ditunjukkan pada Gambar 5 berikut :



Gambar 5. *Tokenizing*

Setelah tokenisasi selesai, dokumen akan dilanjutkan ke tahap *stemming* yaitu proses merubah kata yang telah ditokenisasi menjadi kata dasarnya. Namun tahap *stemming* tidak sering digunakan karena mengakibatkan kerancuan dan menjadi tidak spesifik dalam merepresentasikan arti yang sebenarnya dari kata hasil.

#### 4.2.1.6. *Filtering*

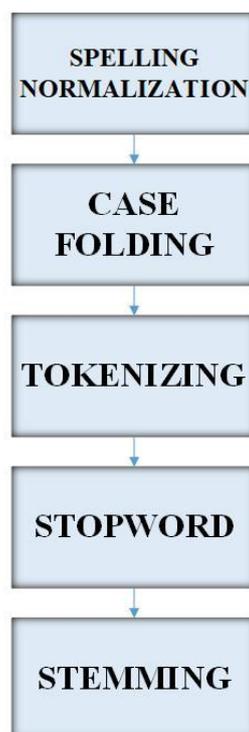
Tahap penyaringan atau *filtering* merupakan tahap dilakukannya pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan tahap untuk menghilangkan katakata yang tidak berpengaruh / tidak informatif namun seringkali muncul dalam dokumen. Kata-kata tersebut seperti kata penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak begitu memiliki arti dalam penentuan kelas topik suatu dokumen.

Contoh proses *filtering* dapat dilihat pada Gambar 6 berikut :



Gambar 6. *Filtering*

*Flowchart* dari proses *text preprocessing* dan *filtering* adalah sebagai berikut:



Gambar 7. *Flowchart Text Mining*

#### 4.2.1.7. Pelabelan Kelas Sentimen

Setelah melalui proses *preprocessing*, selanjutnya akan dilakukan analisis sentimen untuk pelabelan data. Proses pelabelan data dilakukan secara otomatis dengan cara menghitung skor sentimen. Pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata pada sebuah dokumen teks. Semakin sering

sebuah kata muncul pada sebuah dokumen teks, maka bobot kata tersebut semakin besar dan kata tersebut dianggap sebagai kata yang sangat merepresentasikan dokumen teks tersebut. Pada umumnya, analisis sentimen digunakan untuk melakukan klasifikasi (pelabelan) dokumen teks ke dalam tiga kelas sentimen, yaitu sentimen positif, negatif dan netral. Cara menentukan kelas sentimen adalah dengan menghitung skor jumlah kata positif dikurangi skor jumlah kata negatif dalam setiap kalimat ulasan. Kalimat yang memiliki skor  $> 0$  akan diklasifikasikan ke dalam kelas positif, kalimat yang memiliki skor  $= 0$  akan diklasifikasikan ke dalam kelas netral, sedangkan kalimat yang memiliki skor  $< 0$  diklasifikasikan ke dalam kelas negatif. Akan tetapi pada penelitian ini digunakan dua pelabelan kelas sentimen, yaitu sentimen positif dan sentimen negatif.

#### 4.2.1.8. Pelabelan dengan Metode Naive Bayes Classifier

Kasus yang diuji adalah sebagai berikut :

Sebuah file trainer yang dikategorisasikan secara manual ke dalam sentimen positif. Isi file trainer tersebut adalah sebagai berikut : Pandai, cerdas, pas

Sebuah file trainer yang dikategorisasikan secara manual ke dalam sentimen negatif. Isi file trainer tersebut adalah sebagai berikut : Lambat, kurang

Sebuah file tester. Isi file tester tersebut adalah sebagai berikut : Pandai, cerdas, cerdas, pas, lambat, kurang

File tester tersebut berisi kata kata kedua file trainer, namun jumlah kata cerdas lebih banyak (untuk tag positif). Oleh karena itu, keluaran yang diharapkan adalah dapat dikategorisasikan ke dalam dua sentimen, namun dengan nilai probabilitas lebih tinggi untuk sentimen Positif dari pada sentimen Negatif.

Diketahui :

$D$  : sekumpulan kata-kata dalam dokumen tester = {pandai, cerdas, cerdas, pas, lambat, kurang}

$|V|$  : himpunan kata-kata yang diketahui = {pandai, cerdas, pas, lambat, kurang}

Berdasarkan hasil training, prior probability untuk masing-masing sentimen adalah sebagai berikut :

$$P(C \text{ positif}) = 3/5 = 0,6$$

$$P(C \text{ negatif}) = 2/5 = 0,4$$

Hitungan probabilitas masing-masing kata terhadap masing-masing kategori dengan menggunakan smooting laplace untuk menghindari 0.

Perhitungan sentimen positif :

$$P(W \text{ Pandai} | C \text{ Positif}) = \frac{1 + \#words(C \text{ Positif}, W \text{ Pandai})}{|V| + \#words(C \text{ Positif})} = \frac{1 + 1}{5 + 3} = 0,25$$

$$\begin{aligned}
P(W \text{ Cerdas} | C \text{ Positif}) &= \frac{1 + \#words(C \text{ Positif}, W \text{ Cerdas})}{|V| + \#words(C \text{ Positif})} = \frac{1 + 1}{5 + 3} \\
&= 0,25 \\
P(W \text{ Pas} | C \text{ Positif}) &= \frac{1 + \#words(C \text{ Positif}, W \text{ Pas})}{|V| + \#words(C \text{ Positif})} = \frac{1 + 1}{5 + 3} = 0,25 \\
P(W \text{ Lambat} | C \text{ Positif}) &= \frac{1 + \#words(C \text{ Positif}, W \text{ Lambat})}{|V| + \#words(C \text{ Positif})} = \frac{1 + 0}{5 + 3} \\
&= 0,125 \\
P(W \text{ Kurang} | C \text{ Positif}) &= \frac{1 + \#words(C \text{ Positif}, W \text{ Kurang})}{|V| + \#words(C \text{ Positif})} = \frac{1 + 0}{5 + 3} \\
&= 0,125 \\
P(W \text{ Pandai} | \overline{C \text{ Positif}}) &= \frac{1 + \#words(\overline{C \text{ Positif}}, W \text{ Pandai})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 0}{5 + 2} \\
&= 0,142857 \\
P(W \text{ Cerdas} | \overline{C \text{ Positif}}) &= \frac{1 + \#words(\overline{C \text{ Positif}}, W \text{ Cerdas})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 0}{5 + 2} \\
&= 0,142857 \\
P(W \text{ Pas} | \overline{C \text{ Positif}}) &= \frac{1 + \#words(\overline{C \text{ Positif}}, W \text{ Pas})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 0}{5 + 2} \\
&= 0,142857 \\
P(W \text{ Lambat} | \overline{C \text{ Positif}}) &= \frac{1 + \#words(\overline{C \text{ Positif}}, W \text{ Lambat})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 1}{5 + 2} \\
&= 0,2857 \\
P(W \text{ Kurang} | \overline{C \text{ Positif}}) &= \frac{1 + \#words(\overline{C \text{ Positif}}, W \text{ Kurang})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 1}{5 + 2} \\
&= 0,2857
\end{aligned}$$

Dari nilai diatas dapat dihitung probabilitas dokumen terhadap sentimen positif :

$$\begin{aligned}
P(d | C \text{ Positif}) &= \frac{0.25^1}{1!} x \frac{0.25^2}{2!} x \frac{0.25^1}{1!} x \frac{0.125^1}{1!} x \frac{0.1}{1!} x \frac{0.25^1}{1!} = 0,0000305175 \\
P(d | \overline{C \text{ Positif}}) &= \frac{0,142857^1}{1!} x \frac{0,142857^2}{2!} x \frac{0,142857^1}{1!} x \frac{0,2857^1}{1!} x \frac{0,2857^1}{1!} \\
&= 0,000016658
\end{aligned}$$

Hitung probabilitas dokumen tersebut :

$$P(d) = P(d|C \text{ Positif}) \times P(C \text{ Positif}) + P(d|\overline{C \text{ Positif}}) \times P(\overline{C \text{ Positif}})$$

$$0,0000305175 \times 0,6 + 0,000016658 \times 0,4 = 0,0000249737$$

Probabilitas sentimen tersebut terhadap dokumen adalah :

$$P(C \text{ Positif} | d) = \frac{P(d | C \text{ Positif}) \times P(C \text{ Positif})}{P(d)} = \frac{0,0000305175 \times 0,6}{0,0000249737} \\ = 0,73319$$

Diperoleh probabilitas untuk sentimen positif terhadap dokumen tester adalah 73,319%

Perhitungan sentimen negatif :

$$P(W \text{ Pandai} | C \text{ Negatif}) = \frac{1 + \#words(C \text{ Negatif}, W \text{ Pandai})}{|V| + \#words(C \text{ Negatif})} = \frac{1 + 0}{5 + 2} \\ = 0,1428$$

$$P(W \text{ Cerdas} | C \text{ Negatif}) = \frac{1 + \#words(C \text{ Negatif}, W \text{ Cerdas})}{|V| + \#words(C \text{ Negatif})} = \frac{1 + 0}{5 + 2} \\ = 0,1428$$

$$P(W \text{ Pas} | C \text{ Negatif}) = \frac{1 + \#words(C \text{ Negatif}, W \text{ Pas})}{|V| + \#words(C \text{ Negatif})} = \frac{1 + 0}{5 + 2} \\ = 0,1428$$

$$P(W \text{ Lambat} | C \text{ Negatif}) = \frac{1 + \#words(C \text{ Negatif}, W \text{ Lambat})}{|V| + \#words(C \text{ Negatif})} = \frac{1 + 1}{5 + 2} \\ = 0,2857$$

$$P(W \text{ Kurang} | C \text{ Negatif}) = \frac{1 + \#words(C \text{ Negatif}, W \text{ Kurang})}{|V| + \#words(C \text{ Negatif})} = \frac{1 + 1}{5 + 2} \\ = 0,2857$$

$$P(W \text{ Pandai} | \overline{C \text{ Negatif}}) = \frac{1 + \#words(\overline{C \text{ Negatif}}, W \text{ Pandai})}{|V| + \#words(\overline{C \text{ Negatif}})} = \frac{1 + 1}{5 + 3} \\ = 0,25$$

$$(W \text{ Cerdas} | \overline{C \text{ Negatif}}) = \frac{1 + \#words(\overline{C \text{ Negatif}}, W \text{ Cerdas})}{|V| + \#words(\overline{C \text{ Negatif}})} = \frac{1 + 1}{5 + 3} \\ = 0,25$$

$$P(W \text{ Pas} | \overline{C \text{ Negatif}}) = \frac{1 + \#words(\overline{C \text{ Negatif}}, W \text{ Pas})}{|V| + \#words(\overline{C \text{ Negatif}})} = \frac{1 + 1}{5 + 3} \\ = 0,25$$

$$P(W \text{ Lambat} | \overline{C \text{ Negatif}}) = \frac{1 + \#words(\overline{C \text{ Negatif}}, W \text{ Lambat})}{|V| + \#words(\overline{C \text{ Positif}})} = \frac{1 + 0}{5 + 3} \\ = 0,125$$

$$P(W \text{ Kurang} | \overline{C \text{ Negatif}}) = \frac{1 + \#words(\overline{C \text{ Negatif}}, W \text{ Kurang})}{|V| + \#words(\overline{C \text{ Negatif}})} = \frac{1 + 0}{5 + 3} \\ = 0,125$$

Dari nilai diatas dapat dihitung probabilitas dokumen terhadap sentimen Negatif :

$$P(d | C \text{ Negatif}) = \frac{0.1428^1}{1!} x \frac{0.1428^2}{2!} x \frac{0.1428^1}{1!} x \frac{0.2857^1}{1!} x \frac{0.2857^1}{1!} \\ = 0,0000166447$$

$$P(d | \overline{C \text{ Negatif}}) = \frac{0,25^1}{1!} x \frac{0,25^2}{2!} x \frac{0,25^1}{1!} x \frac{0,125^1}{1!} x \frac{0,125^1}{1!} \\ = 0,0000000305175$$

Hitung probabilitas dokumen tersebut :

$$P(d) = P(d|C \text{ Negatif}) x P(C \text{ Negatif}) + P(d|\overline{C \text{ Negatif}}) x P(\overline{C \text{ Negatif}})$$

$$0,0000166447 x 0,4 + 0,0000305175 X 0,6 = 0,00002493$$

Probabilitas sentimen tersebut terhadap dokumen adalah :

$$P(C \text{ Negatif}|d) = \frac{P(d|C \text{ Negatif})x P(C \text{ Negatif})}{P(d)} = \frac{0,0000166447 x 0,4}{0,00002493} \\ = 0,26706$$

Diperoleh probabilitas untuk sentimen positif terhadap dokumen tester adalah 26,706%

Untuk sentimen positif adalah = 73,319 %

Untuk sentimen negatif adalah = 26,706%

Sehingga dapat disimpulkan bahwa file tester masuk ke dalam sentimen positif.

#### 4.2.1.9 Pelabelan dengan *Support Vector Machine*

Untuk penggunaan metode *Support Vector Machine* didasarkan pada label dan kata-kata yang ada di dalam teks. Misal terdapat tiga buah dokumen yang diambil dari kelas positif yaitu:

Dokumen 1 ( $d_1$ ) = Pandai, cerdas, baik, pas, lambat, kurang  
 Dokumen 2 ( $d_2$ ) = Pandai, kurang, lambat, buruk  
 Dokumen 3 ( $d_3$ ) = Lambat, cerdas, baik, pas, lambat, kurang

Tahap Pertama merupakan proses penyajian data dokumen dalam bentuk tabel kemudian dilakukan normalisasi matriks.

**Tabel 3. Tabel Data Awal**

Dokumen	+	-	0
D1	4	2	0
D2	1	3	0
D3	3	3	1

Tahap normalisasi data :

$$X_{11} = \frac{4 - 0}{4 - 0} = 1$$

$$X_{12} = \frac{2 - 0}{4 - 0} = 0,5$$

$$X_{13} = X_{23} = 0$$

$$X_{21} = X_{33} = \frac{1 - 0}{4 - 0} = 0,25$$

$$X_{22} = X_{31} = X_{32} = \frac{3 - 0}{4 - 0} = 0,75$$

**Tabel 4. Tabel Normalisasi Data**

Dokumen	+	-	0
D1	1	0,5	0
D2	0,25	0,75	0
D3	0,75	0,75	0,25

Tahap Heusian (Determinan dan Invers) :

$$M_{11} = \begin{bmatrix} 0,75 & 0 \\ 0,75 & 0,25 \end{bmatrix} = 0,1875 - 0 = 0,1875$$

$$M_{21} = \begin{bmatrix} 0,5 & 0 \\ 0,75 & 0,25 \end{bmatrix} = 0,125 - 0 = 0,125$$

$$M_{31} = \begin{bmatrix} 0,5 & 0 \\ 0,75 & 0 \end{bmatrix} = 0$$

$$M_{12} = \begin{bmatrix} 0,25 & 0 \\ 0,75 & 0,25 \end{bmatrix} = 0,0625 - 0 = 0,0625$$

$$M_{22} = \begin{bmatrix} 1 & 0 \\ 0,75 & 0,25 \end{bmatrix} = 0,25$$

$$M_{32} = \begin{bmatrix} 1 & 0 \\ 0,25 & 0 \end{bmatrix} = 0$$

$$M_{31} = \begin{bmatrix} 0,5 & 0 \\ 0,75 & 0 \end{bmatrix} = 0$$

$$M_{32} = \begin{bmatrix} 1 & 0 \\ 0,25 & 0 \end{bmatrix} = 0$$

$$M_{33} = \begin{bmatrix} 1 & 0,5 \\ 0,25 & 0,75 \end{bmatrix} = 0,75 - 0,125 = 0,625$$

**Tabel 5. Tabel Heusian**

<b>Dokumen</b>	<b>+</b>	<b>-</b>	<b>0</b>
<b>D1</b>	0,1875	0,125	0
<b>D2</b>	0,0625	0,25	0
<b>D3</b>	0	0	0,625

Tahap Sequential Matriks :

$$D_{11} = 0,035 \left( 2,718 \left( \frac{1}{1 \times 0,49} \right) \right) + 18,66 = 18,854$$

$$D_{12} = 0,25 \left( 2,718 \left( \frac{0,5}{1 \times 0,49} \right) \right) + 18,66 = 19,353$$

$$D_{13} = 0 \left( 2,718 \left( \frac{0}{1 \times 0,49} \right) \right) + 18,66 = 18,66$$

$$D_{21} = 0,0625 \left( 2,718 \left( \frac{0,25}{1 \times 0,49} \right) \right) + 18,66 = 18,746$$

$$D_{22} = 0,562 \left( 02,718 \left( \frac{0,75}{1 \times 0,49} \right) \right) + 18,66 = 19,43$$

$$D_{23} = 0 \left( 2,718 \left( \frac{0}{1 \times 0,49} \right) \right) + 18,66 = 18,66$$

$$D_{31} = 0,562 \left( 02,718 \left( \frac{0,75}{1 \times 0,49} \right) \right) + 18,66 = 19,43$$

$$D_{32} = 0,562 \left( 02,718 \left( \frac{0,75}{1 \times 0,49} \right) \right) + 18,66 = 19,43$$

$$D_{33} = 0,0625 \left( 2,718 \left( \frac{0,25}{1 \times 0,49} \right) \right) + 18,66 = 18,746$$

Tahap klasifikasi dengan kondisi jika  $D < 0$  maka termasuk sentimen positif. Jika  $D > 0$  maka termasuk sentimen negatif. Jika  $D = 0$  maka termasuk sentimen netral.

**Tabel 6. Tabel Pelabelan Sentimen**

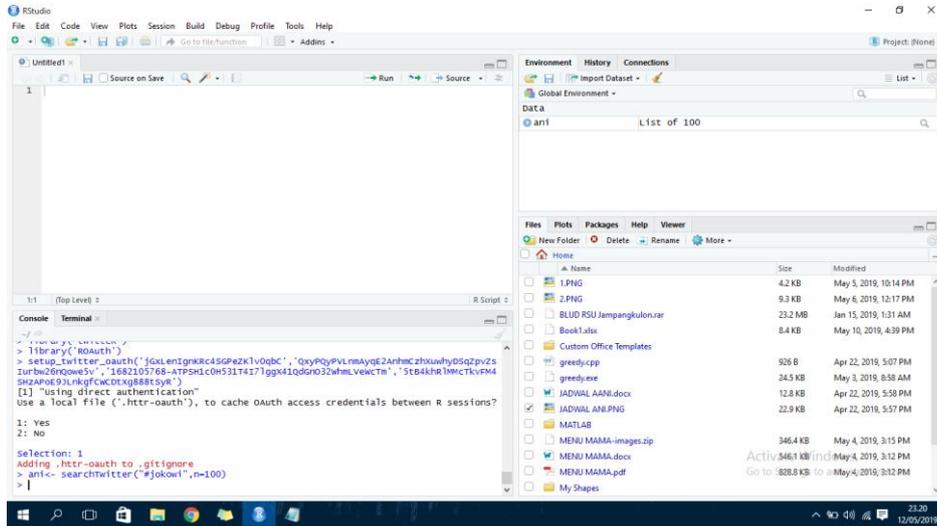
Dokumen	+	-	Sentimen
<b>D1</b>	19,353	19,353	Positif
<b>D2</b>	18,746	19,43	Negatif
<b>D3</b>	19,43	19,43	Netral

### 4.3. Implementation (Implementasi)

Pada tahap ini, pertama kali dikembangkan di program kecil yang disebut unit, yang terintegrasi dalam tahap selanjutnya. Setiap unit dikembangkan dan diuji untuk fungsionalitas yang disebut sebagai *unit testing*.

### 4.3.1. Proses Scraping

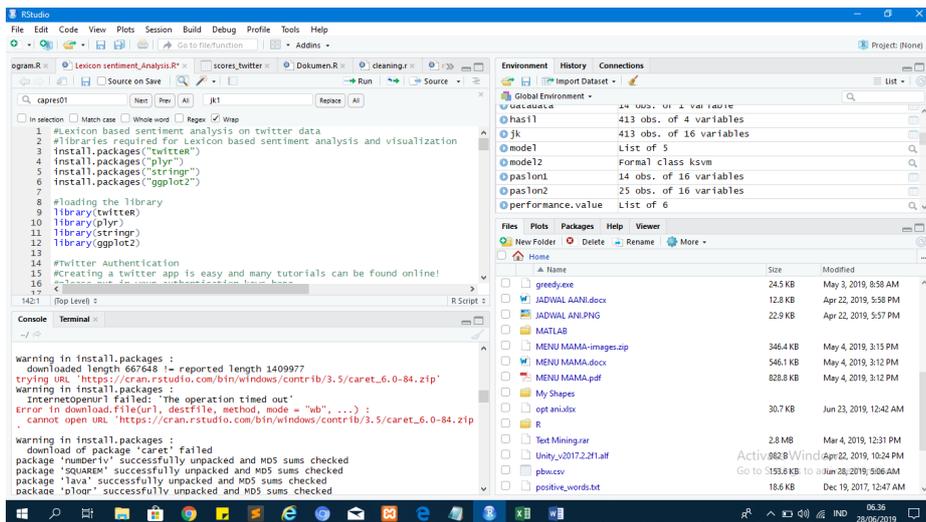
Dibawah ini merupakan proses implementasi scraping untuk pengambilan data twitter sebanyak 600 data.



Gambar 8. Proses Scraping

### 4.3.2. Proses Text Mining

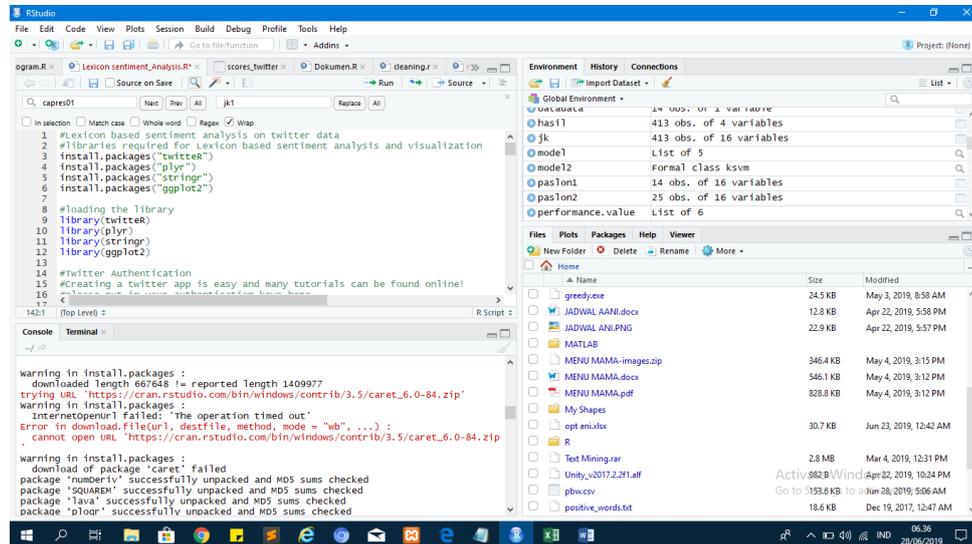
Pada gambar 9 merupakan proses *text mining* dengan urutan *tolowercase*, *rokenizing*, *stopword*, dan *stemming*



Gambar 9. Proses Pengolahan Data dan Kata

### 4.3.4. Proses Klasifikasi

Proses klasifikasi merupakan pembobotan atau pembagian kata positif dan negatif berdasarkan metode *SVM* dan *MBC*



Gambar 10. Proses Klasifikasi dengan 2 Metode

### 4.4. Integration & Testing (Uji Coba)

Seluruh *unit* yang dikembangkan dalam tahap implementasi diintegrasikan ke dalam sistem setelah pengujian yang dilakukan masing-masing *unit*. Setelah integrasi seluruh sistem diuji untuk mengecek setiap kegagalan maupun kesalahan yaitu dengan menggunakan *confusion Matriks*

# BAB V

## HASIL DAN PEMBAHASAN

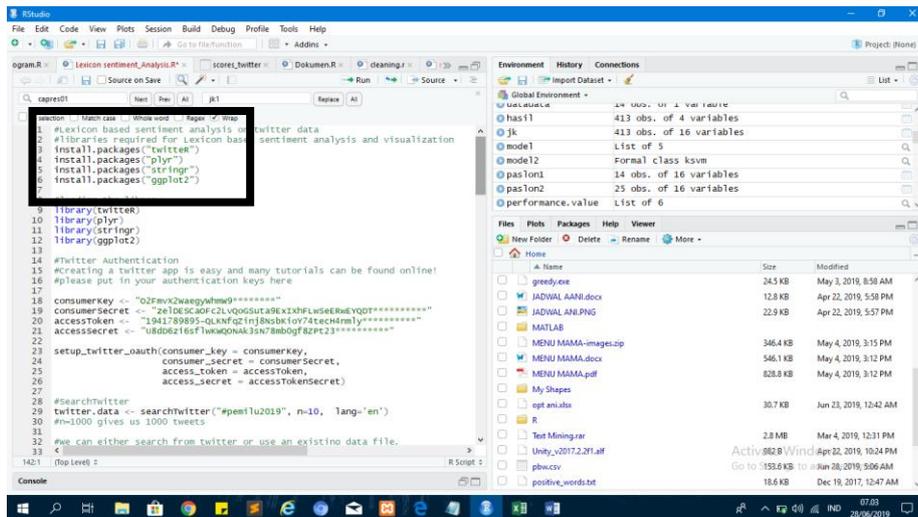
### 5.1. Hasil

Pada tahap perancangan dan implementasi yang sebelumnya telah dijelaskan mengenai rancangan. Berikut ini adalah hasil beserta pembahasan program:

#### 5.1.1. Instalasi Package

Pada proses ini, dilakukan instalasi *package* untuk menyediakan tempat yang dapat digunakan untuk mengakses *library* agar program dapat berjalan dengan program sebagai berikut :

```
install.packages('httr'), install.packages('twitteR'), install.packages('ROAuth'),  
install.packages("tm"), install.packages("SnowballC"), install.packages("plyr"),  
install.packages("stringr"), install.packages("ggplot2")
```

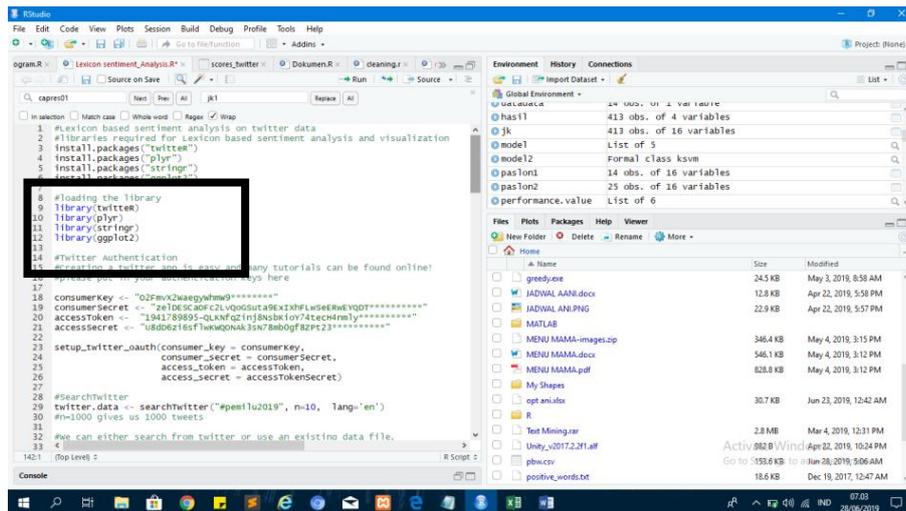


Gambar 11. Proses Instalasi Package

#### 5.1.2. Pemanggilan Library

Pada proses ini, dilakukan pemanggilan *library* untuk menjalankan program program untuk melakukan pengolahan kata serta pengelompokan kata dengan program sebagai berikut :

```
library('httr'), library('twitteR'), library('ROAuth'), library(tm), library(SnowballC),  
library(twitteR), library(plyr), library(stringr), library(ggplot2)
```



Gambar 12. Proses Pemanggilan Library

### 5.1.3. Proses Scraping

Pada proses ini, dilakukan pengambilan data pada twitter dengan menggunakan API berdasarkan hastag yang telah ditentukan untuk keyword scraping dengan program sebagai berikut :

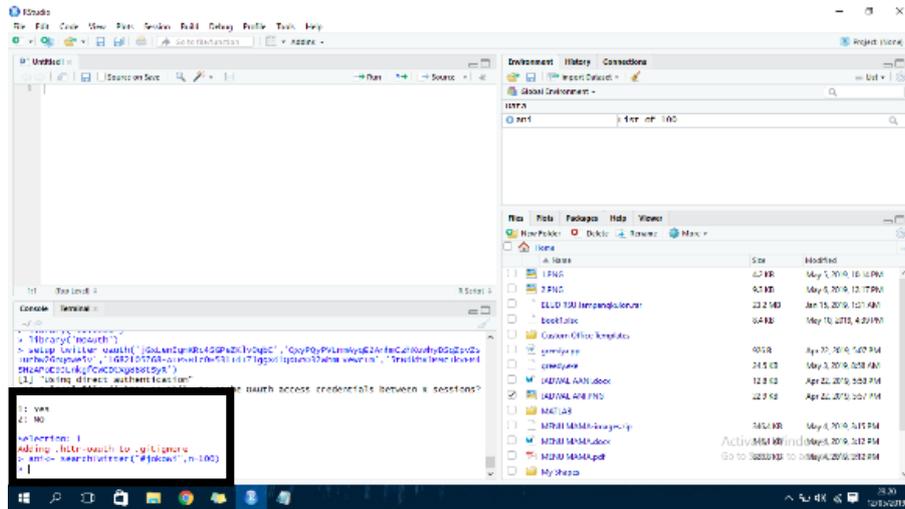
```

setup_twitter_oauth('jGxLenIgnKRc4SGPeZKlv0qbC','QxyPQyPVLnmAyyqE2AnhmC
zhXuwhyDSqZpvZslurbw26nQowe5v','1682105768-
ATPSH1c0H531T4I7lggX41QdGnO32WhmLVeWcTm','5tB4khRIMMcTkVFM4SHzA
PoE9JLnlkgfCWCDtXg888tSyR')
tweets <- searchTwitter("#prabowo",n=250)

tweets <- twListToDF(tweets)

> write.csv(tweets, "pemilu2019.csv", row.names = FALSE)

```

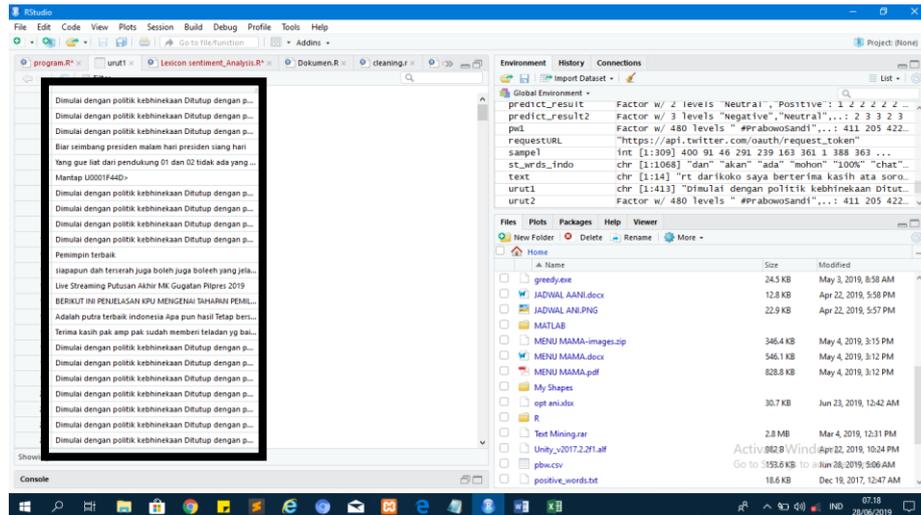


Gambar 13. Proses Scraping Data

#### 5.1.4. Proses Text Mining

Pada proses ini, dilakukan pembersihan data dimulai dari menghilangkan beberapa simbol dan karakter tidak penting serta menghilangkan link yang tertaut dengan program sebagai berikut :

```
urut1 <- sapply(urut1, function(row) iconv(row, "latin1", "ASCII", sub=""))
urut1 <- gsub("@\\w+", "", urut1)
urut1 <- gsub("#\\w+", "", urut1)
urut1 <- gsub("RT\\w+", "", urut1)
urut1 <- gsub("http.*", "", urut1)
urut1 <- gsub("RT", "", urut1)
urut1 <- sub("[.-]//[[:punct:]]", "\\1", urut1)
urut1 <- sub("[']//[[:punct:]]", "\\1", urut1)
```

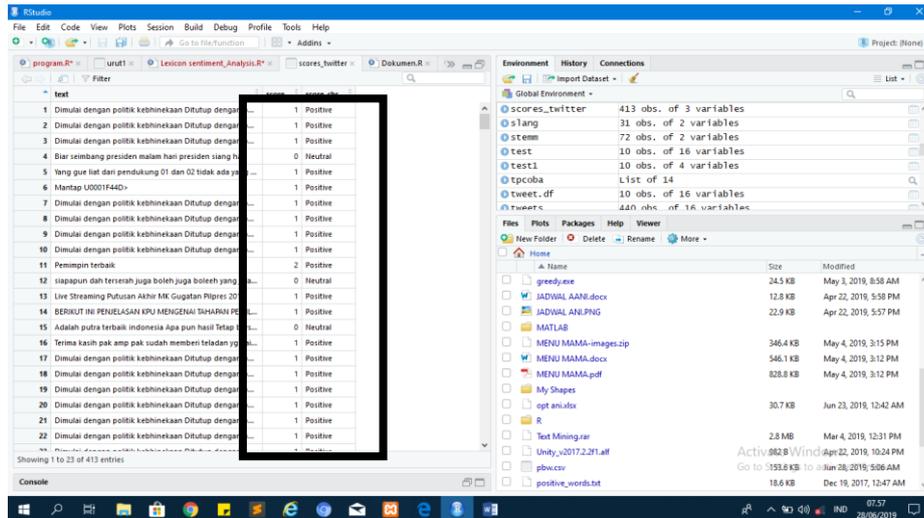


Gambar 14. Proses Text Mining

### 5.1.5. Proses Pembacaan Corpus dan Pelabelan Score

Pada proses ini, dilakukan pembacaan corpus negatif dan positif serta membeli label pada kalimat atau text yang memiliki label sentimen sesuai dengan corpus dengan program sebagai berikut :

```
pos <- readLines("positif.txt")
neg <- readLines("negatif.txt")
scores_twitter <- score.sentiment(urut1, pos, neg, .progress='text')
scores_twitter$score_chr <- gsub("^0$", "Neutral", scores_twitter$score_chr)
scores_twitter$score_chr <- gsub("^1$/^2$/^3$/^4$", "Positive",
scores_twitter$score_chr)
scores_twitter$score_chr <- gsub("^5$/^6$/^7$/^8$/^9$/^10$/^11$/^12$", "Very
Positive", scores_twitter$score_chr)
scores_twitter$score_chr <- gsub("^-1$/^-2$/^-3$/^-4$", "Negative",
scores_twitter$score_chr)
scores_twitter$score_chr <- gsub("^-5$/^-6$/^-7$/^-8$/^-9$/^-10$/^-11$/^-12$",
"Very Negative", scores_twitter$score_chr)
```

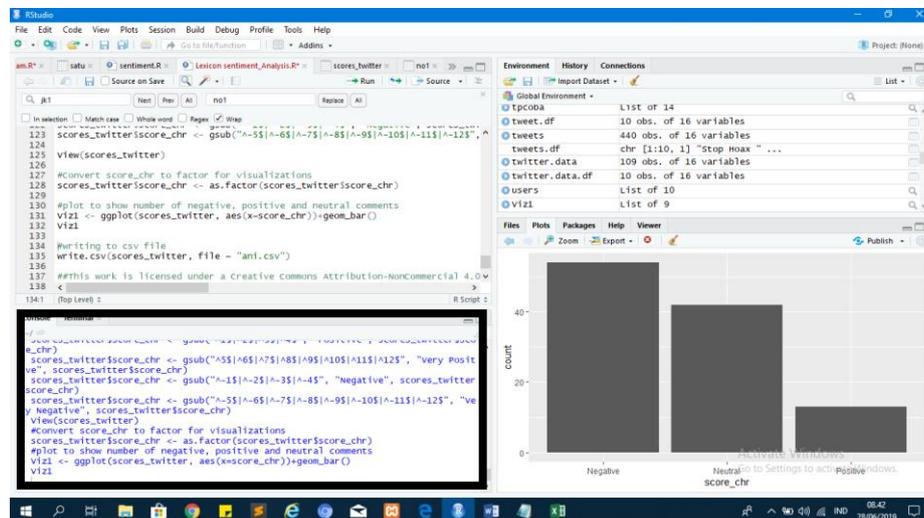


Gambar 15. Proses Pembacaan Corpus dan Pelabelan Score

### 5.1.6. Proses Pembuatan Grafik

Pada proses ini, dilakukan penggambaran informasi data ke dalam sebuah grafik batang dengan program sebagai berikut :

```
Viz1 <- ggplot(scores_twitter, aes(x=score_chr))+geom_bar()
```



Gambar 16. Proses Pembuatan Grafik

### 5.1.7. Proses Klasifikasi dengan Metode NBC

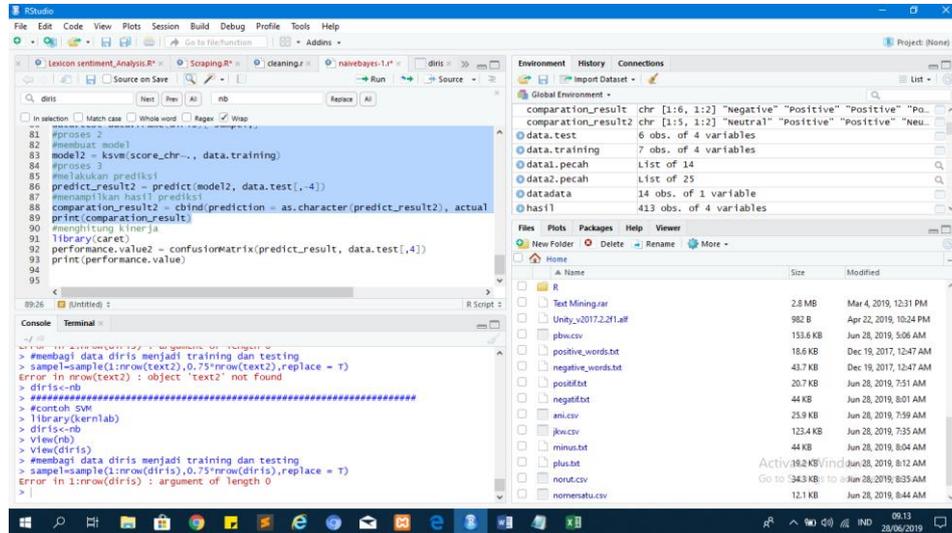
Pada proses ini, dilakukan proses pengelompokan data berdasarkan corpus dengan menggunakan metode NBC dengan program sebagai berikut :



```
predict_result2 = predict(model2, data.test[,-4])
```

```
comparation_result2 = cbind(prediction = as.character(predict_result2), actual =  
as.character(data.test[,4]))
```

```
print(comparation_result)
```



Gambar 18. Proses Klasifikasi dengan Metode SVM

## 5.2. Pembahasan

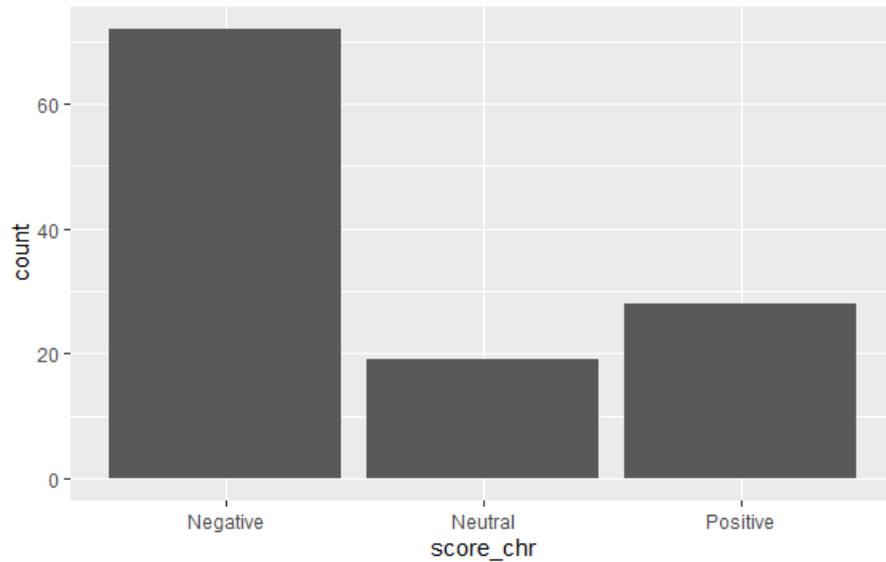
Berdasarkan program yang telah diimplementasikan terdapat beberapa langkah yang harus dilakukan untuk mengklasifikasi 3 variabel berdasarkan 2 metode dalam dua bahasa yaitu bahasa indonesia dan bahasa inggris. Dalam tahap implmentasi terdapat kendala terutama pada bagian *scraping program*, karena *scraping* didasarkan pada *keyword* berupa *hashtag* sehingga semua *tweet* yang menggunakan *hashtag* tersebut akan tersaring. Jadi diperlukannya filtering tahap pertama secara manual dengan menggunakan aplikasi *MS. Excel*.

### 5.2.1. Hasil Analisis Sentimen

Berdasarkan hasil analisis sentimen pada 3 variabel yaitu negatif, netral, dan positif maka diperoleh hasil sebagai berikut :

#### 5.2.1.1. Hasil Analisis Sentimen berbahasa indonesia Calon Presiden Nomer Urut 1

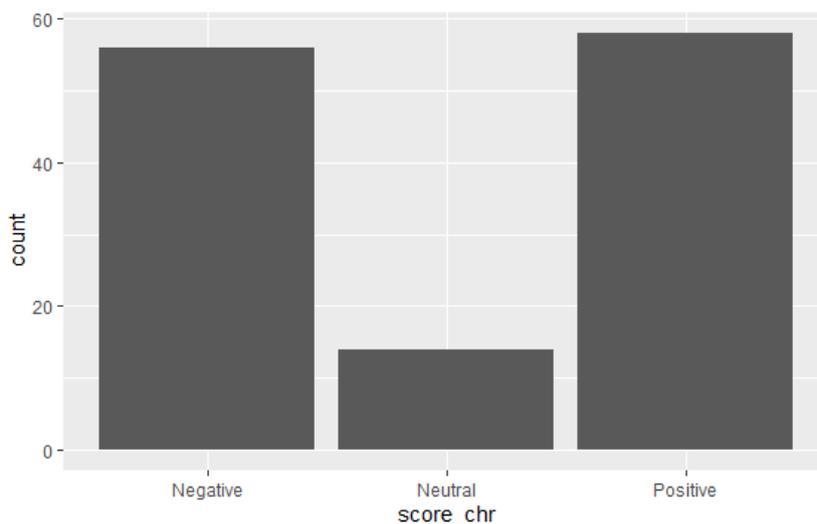
Berdasarkan gambar 19 dapat diketahui bahwa pasangan capres nomer urut 1 memperoleh sentimen negatif lebih banyak dibandingkan dengan sentimen netral dan sentimen positif. Untuk sentiment negatif sebanyak 60,8%, sentiment netral sebanyak 15,8% dan sentimen positif sebanyak 23,4%



Gambar 19. Grafik Analisis Sentimen berbahasa indonesia Capres No Urut 1

#### 5.2.1.2. Hasil Analisis Sentimen berbahasa indonesia Calon Presiden Nomer Urut 2

Berdasarkan gambar 20 dapat diketahui bahwa pasangan capres nomer urut 2 memperoleh sentimen positif lebih banyak dibandingkan dengan sentimen netral dan sentimen negatif. Untuk sentiment negatif sebanyak 42,9 %, sentiment netral sebanyak 10,9 % dan sentimen positif sebanyak 46,2%

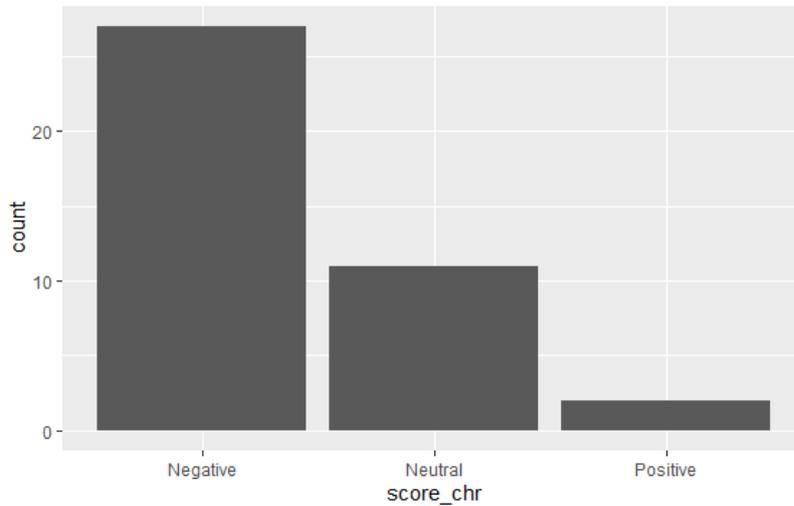


Gambar 20. Grafik Analisis Sentimen berbahasa Indonesia Capres No Urut 2

#### 5.2.1.2. Hasil Analisis Sentimen berbahasa inggris Calon Presiden Nomer Urut 1

Berdasarkan gambar 21 dapat diketahui bahwa pasangan capres nomer urut 1 memperoleh sentimen negatif lebih banyak dibandingkan dengan sentimen netral dan

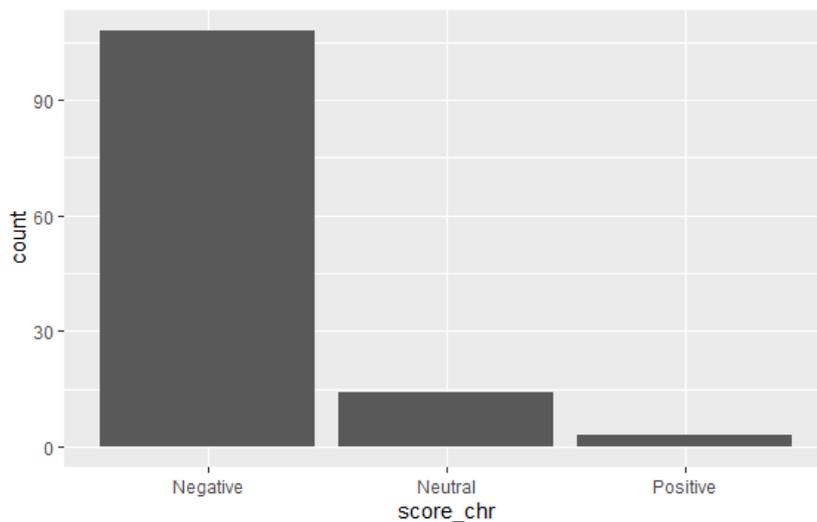
sentimen positif. Untuk sentiment negatif sebanyak 72,5 %, sentiment netral sebanyak 22,5 % dan sentimen positif sebanyak 5%.



Gambar 21. Grafik Analisis Sentimen berbahasa Inggris Capres No Urut 1

#### 5.2.1.2. Hasil Analisis Sentimen berbahasa inggris Calon Presiden Nomer Urut 2

Berdasarkan gambar 22 dapat diketahui bahwa pasangan capres nomer urut 2 memperoleh sentimen negative lebih banyak dibandingkan dengan sentimen netral dan sentimen positif. Untuk sentiment negatif sebanyak 86,4 %, sentiment netral sebanyak 10,4 % dan sentimen positif sebanyak 3,2%.



Gambar 22. Grafik Analisis Sentimen berbahasa Inggris Capres No Urut 2

**Tabel 7. Tabel Hasil Analisis Sentimen**

Capres	Sentimen	Inggris	Indonesia	Rata-Rata
01	Negatif	72,5%	60,8%	68,15%
	Positif	5%	23,4%	14,2%
	Netral	22,5%	15,8%	19,5%
02	Negatif	86,4%	42,9%	64,65
	Positif	3,2%	46,2%	24,7%
	Netral	10,4%	10,9%	10,65%

Untuk mengetahui apakah aplikasi memenuhi syarat maka wajib dilakukan tahap uji coba. Uji coba sistem merupakan pengujian terhadap program yang telah dibuat, dengan uji coba ini dapat diketahui nilai akurasi berdasarkan kedua metode dan kedua bahasa.

### 5.2.2. Hasil Akurasi *Confusion Martix*

Tabel dibawah ini merupakan Tabel hasil akurasi untuk sentiment berbahasa inggris dan berbahasa Indonesia dengan menggunakan 2 metode.

**Tabel 8. Tabel Akurasi**

Capres	Bahasa	Metode	Akurasi	Rata-rata akurasi
01	Indonesia	NBC	100%	94%
	Inggris	NBC	88,24%	
	Indonesia	SVM	50,80%	72%
	Inggris	SVM	94,12%	
02	Indonesia	NBC	98,39%	99%
	Inggris	NBC	100%	
	Indonesia	SVM	35,57%	65%
	Inggris	SVM	94,55%	

Berdasarkan tabel akurasi berikut dapat diketahui bahwa metode NBC lebih cocok digunakan untuk kasus analisis sentimen masyarakat terhadap peserta pemilu 2019 pada media social twitter dikarenakan nilai akurasinya yang mencapai 94% dan 99% dibandingkan dengan metode SVM yang memiliki nilai akurasi 72% dan 65% .

## **BAB VI**

### **KESIMPULAN DAN SARAN**

#### **6.1. Kesimpulan**

Dari hasil penelitian dan pembahasan yang dilakukan, maka dapat disimpulkan bahwa penarikan data dengan menggunakan media social twitter telah berhasil dilakukan, dengan melakukan Authentication terlebih dahulu untuk mengakses Twitter API. Proses Crawling data bersifat real time, 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. Analisis sentiment dengan menggunakan text mining yang didalamnya terdapat tahapan preprocessing dan ekstraksi fitur untuk menghasilkan tweet kotor menjadi tweet bersih. Algoritma *Naïve Bayes Classifier* merupakan algoritma yang tepat untuk kasus analisis sentimen masyarakat media sosial *Twitter* dengan akurasi 94% dan 99% sedangkan pada metode Support Vector Machine hanya memperoleh akurasi 72% dan 65%.

#### **6.2. Saran**

Dari kesimpulan yang telah diuraikan maka terdapat saran yang perlu disampaikan untuk penelitian selanjutnya yaitu penelitian selanjutnya dapat menggunakan metode lain untuk mengklasifikasi dan menggunakan banyak data latih sehingga hasil yang didapat semakin akurat.

## DAFTAR PUSTAKA

- Ade, Maya & Romi** . 2014. deteksian Kesamaan pada Dokumen Teks Menggunakan Kombinasi Algoritma Enhanced Confix Stripping dan Algoritma Winnowing. Teknologi. Universitas Sumatera Utara.
- Agung P, Rini I & Arie N.** 2017. Sentiment Analysis Tokoh Politik Pada Twitter. Teknologi. Kediri.
- Boy Utomo M.** 2014. Analisis Sentimen Pada Twitter Menggunakan Text Mining. Skripsi. Universitas Sumatera Utara. Medan.
- Feldman, R & Sanger, J.** 2007. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data.* Cambridge University Press, New York.
- Khoiril H, Zulfa & Azzah F.** 2018. Pemilu Presiden 2019: Antara Kontestasi Politik dan Persaingan Pemicu Perpecahan Bangsa. Hukum. 4(3) : 547-562.
- Liu, Bing.** 2012. *Sentiment Analysis And Opinion Mining,* Chicago.
- Novantirani, Anita.** 2015. Analisis Sentimen Pada Twitter Mengenai Penggunaan Transportasi Umum Darat Dengan Kota dengan Metode Support Vector Machine. Teknologi. 2(1) : 1177.
- Rossiter, D. G.** 2009. *Introduction to the R Project for Statistical Computing for use at ITC.* University of Twente, Enschede.
- Tala, Fadillah Z.** 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia.* Institute for Logic, Language and Computation Universiteit van Amsterdam, Netherlands.
- Zhang, T.** 2005. *Text Mining : Predictive Methods fo Analyzing Unstructured Information.* Springer, New York.

# LAMPIRAN

Lampiran 1. Kartu Bimbingan

**Kartu Bimbingan Mahasiswa  
Program Studi Ilmu Komputer FMIPA - UNPAK**

Nama Mahasiswa : Fany Syafira Nurul Kilah  
 NPM : 065119125  
 Judul Skripsi : Analisis Sentimen Masyarakat Terhadap Calon Presiden 2019 pada media sosial Twitter Menggunakan Metode Naive Bayes Classifier  
 Pembimbing I : Prihaslati Harsoni, M.Si  
 Pembimbing II : Irma Angraeni, M.kom

No.	Hari, tanggal	Catatan	Tanda Tangan	
			Pemb. I	Pemb. II
1.		Penentuan topic		
2.				
3.	22/03/2019	Penentuan topic		
4.				
5.	27/03/2019	Pengisian sub bab		
6.		Lampiran bab IV		
7.	01-07-2019	Revisi Laporan		
8.				
9.	03-07-2019	Revisi Laporan		
10.				
11.		Revisi Program		
12.				
13.	17-07-2019	Revisi Program dan Laporan		
14.				
15.		Penyempurnaan Abg.		
16.				
17.				
18.				

Bogor, ..... 20.....  
 Program Studi Ilmu Komputer  
 Fakultas MIPA - UNPAK  
 Ketua  
  
 Prihaslati Harsoni, M.Si

Lampiran 2. Kartu Hadir Seminar

\* Cetak dengan kertas pink

**Kartu Daftar Hadir Seminar**  
**Program Studi Ilmu Komputer**

Nama : Fany Syafira N.k.....  
NPM : 065115125.....

No.	Hari / Tanggal	Seminar		Judul Seminar yang diikuti	Paraf
		Proposal	Hasil		
1	Jumat, 17-Mei-2019	✓		Model inkubator Bayi Menggunakan Sensor DHT-22 Berbasis Mikrokontroler Donemcu TSP 8266	
2	Jumat, 17-Mei-2019	✓		Model Peringatan Dini Dan Notifikasi Tanah Longsor Menggunakan Sensor Gerak Dan Inter of Things Berbasis Mikrokontroler	
3	Rabu 03-Juli-2019		✓	Optimasi Performansi server Dengan menerapkan Load Balancing Dengan Algoritma Weighted Round Robin	

Lampiran 3. Surat Keterangan Maju Sidang Skripsi

\* Cetak dengan kertas pink

**Kartu ACC Maju Seminar**

Nama : Fany Syafira Nurul Kilah  
NPM : 065115125  
Judul Skripsi : Analisis Sentimen Masyarakat Terhadap Peserta Pemilu pada Media Sosial Twitter Menggunakan Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine  
Pembimbing Utama : Prihastuti Harsani, M.Si  
Pembimbing Pendamping : Irma Anggraeni, M.Kom  
Seminar :  Proposal  Hasil  Skripsi

Tanggal Persetujuan Maju Seminar : 16 Juli 2019

Pembimbing Utama



(Prihastuti Harsani, M.Si)

Pembimbing Pendamping



(Irma Anggraeni, M.Kom)

## Lampiran 4. Korpus Kata

