ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019

SKRIPSI

Disusun Oleh:

Rafiqa Cahyani

H76215026

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2019

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini,

Nama	: Rafiqa Cahyani
NIM	: H76215026
Program Studi	: Sistem Informasi
Angkatan	: 2015

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019". Apabila suatu saat nanti terbukti saya melakukantindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian peryataan keaslian ini saya buat dengan sebenar benarnya.

Surabaya, 03 Oktober 2019

Yang menyatakan,

METERAI TEMPEL

6000

ENAM RIBU RUPIAH

(Rafiqa Cahyani)

NIM : H76215026

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh:

NAMA : Rafiqa Cahyani

NIM : H76215026

JUDUL : ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya,30 September 2019

Dosen Pembimbing 1	DosenPembimbing 2
(Indri Sudanawati Rozas, M. Kom)	(Nita Yalina, S.Kom., M.MT)
NIP 198207212014032001	NIP 198702082014032003

Ketua Program Studi,
Sistem Informasi

(Muhammad Andik Izzuddin, M.T)
NIP 198403072014031001

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Rafiqa Cahyani ini telah dipertahankan

di depan tim penguji skripsi

di Surabaya, 03 Oktober 2019

Mengesahkan,

Dewan Penguji

Penguji I	Penguji II
(Indri Sudanawati Rozas, M. Kom)	(Nita Yalina, S.Kom., M.MT)
NIP 198207212014032001	NIP 198702082014032003
Penguji III	Penguji IV
(Nurissaidah Ulinnuha, M. Kom)	(Mujib Ridwan, S.Kom., M.T)
NIP 199011022014032004	NIP 198604272014031004

Mengetahui

Dekan Fakultas Sains dan Teknologi

UIN Sunan Ampel Surabaya

Dr. Evi Purwati, M.Ag.

NIP 196512211990022001

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama	:	Rafiqa Cahyani
NIM	:	H76215026
Fakultas/Jurusan	:	Sains dan Teknologi / Sistem Informasi
E-mail address	:	rafikacahyani1712@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

▬ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (……)

yang berjudul :

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 14 Oktober 2019

Penulis

(Rafiqa Cahyani)

ABSTRAK

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH PUBLIK PESERTA PILPRES 2019

Oleh:

Rafiqa Cahyani

Di Indonesia twitter menjadi salah satu situs *microblogging* yang sangat digemari oleh masyarakat. Masyarakat sudah terbiasa dalam menyampaikan pendapat atau opininya dalam media sosial termasuk dalam twitter, hal ini digunakan dalam menanggapi suatu kejadian atau peristiwa yang terjadi di lingkungan mereka. Pada penelitian ini dilakakan berbagai proses yaitu *crawling* data dari twitter, pre prosesing data teks, *Naïve Bayes Classifier* untuk probabilitas teks terhadap sentimen, dan *Simple Wighted Product Model* untuk perangkingan sentimen. Hasil yang diperoleh dari penelitian ini adalah peforma sistem untuk setiap tokoh publik dengan menggunakan tiga macam skrenario pembagian data dan dua macam pengambilan data secara urut dan acak. Yang pertama 75% data training dan 25% data testing, yang kedua yaitu 80% dat training dan 20% data testing dan yang terakhir yaitu 90% data training dan 10% data testing. Dari hasil percobaan tersebut dapat diperoleh hasil akurasi terbaik yaitu menggunakan proporsi ketiga dengan pengambilan data secara acak dengan mendapatkan nilai rata rata akurasi sebesar 99,75%. Sedangkan untuk perangkingan sentimen terhadap masing masing tokoh publik dapat di ketahui yang memperoleh peringkat pertama yaitu KH Ma'ruf Amin , yang kedua yaitu Sandiaga Salahudin Uno, yang ketiga yaitu Joko Widodo, dan yang terakhir yaitu Prabowo Subianto.

Kata Kunci: Twitter, Tokoh Publik, *Naïve Bayes Clasifier*, *Simple Wighted Product Model*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

ANALYSIS OF SENTIMENTS IN TWITTER SOCIAL MEDIA ON PUBLIC FIGURE OF PILPRES 2019 PARTICIPANTS

By:

Rafiqa Cahyani

In Indonesia, Twitter has become a microblogging site that is very popular with the public. People are used to expressing their opinions or opinions on social media including on Twitter, this is used in response to an event or events that occur in their environment. In this study various processes were carried out, namely crawling data from Twitter, pre-processing text data, Naïve Bayes Classifier for text probability to sentiment, and Simple Wighted Product Model for ranking sentiments. The results obtained from this study are the performance of the system for each public figure using three types of data sharing scenarios and two kinds of data collection in sequence and random. The first 75% training data and 25% testing data, the second is 80% training data and 20% testing data and the last is 90% training data and 10% testing data. From the results of these experiments can be obtained the best accuracy results using the third proportion with random data collection by getting an average accuracy value of 99.75%. As for ranking sentiments for each public figure, it can be seen that the first rank is KH Ma'ruf Amin, the second is Sandiaga Salahudin Uno, the third is Joko Widodo, and the last is Prabowo Subianto.

Keywords: *Twitter, Public Figure, Naïve Bayes Clasifier, Simple Wighted Product Model*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Daftar Isi

HALAMAN JUDUL i
PERNYATAAN KEASLIAN ii
LEMBAR PERSETUJUAN PEMBIMBING iii
PENGESAHAN TIM PENGUJI SKRIPSI iv
PERNYATAAN PERSETUJUAN PUBLIKASI v
ABSTRAK vi
ABSTRACT vi
Daftar Isi viii
Daftar Gambar ix
Daftar Tabel x
Daftar Lampiran xi
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
BAB 2 TINJAUAN PUSTAKA 7
2.1 Penelitian Terdahulu 7
2.2 Landasan Teori 9
2.2.1 *Machine Learning* 9
2.2.2. Analisis Sentimen 10
2.2.3. *Text Mining* 11
2.2.4 *Naïve Bayes Classifier* 13
2.2.5. *Simple Weighted Product Model* 15
2.2.6. Twitter 17
2.2.7. Kamus Sentiword 18
2.2.8. Pre-Proses Teks 18

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.9. Evaluasi Sistem Klasifikasi 22
2.2.10. *R Programing* 24
2.3 Integrasi Keilmuan.....25
BAB 3 METODOLOGI PENELITIAN.....28
3.1 Jenis Penelitian 28
3.2 Jenis dan Sumber Data.....28
3.3 Metodologi.....29
3.3.1 C*rawling* Data.....30
3.3.2 *Prepocessing*.....32
3.3.3 Pembobotan Kata.....36
3.3.4 Naïve Bayes untuk Probabilitas Sentimen.....37
3.3.5. Proses Metode Simple Weighted Product Model.....39
3.3.6 Analisis Hasil.....40
BAB 4 HASIL DAN PEMBAHASAN.....42
4.1. *Crawling* Data Twitter.....42
4.2. *Prepocessing*.....46
4.3 Manual dan *Rulebase Clasification*.....47
4.4. Pembobotan Kata.....47
4.4.1 Contoh Pembobotan Kata untuk Joko Widodo 48
4.5. *Naïve Bayes* Untuk Probabilitas Teks Terhadap Sentimen 51
4.5.1 Skenario Proporsi Pertama 75 % Training 25 % Testing 52
4.5.2 Skenario Proporsi Kedua 80 % Training 20 % Testing.....57
4.5.3 Skenario Proporsi Ketiga 90 % Training 10 % Testing.....62
4.6. Hasil Analisis Rata-Rata Akurasi Pada Setiap Proporsi.....67
4.7. Analisis Prosentasi Terhadap Jumlah Sentimen Positif dan Negatif.....69
4.8. Perhitungan Menggunakan *Metode Simple Weighted Product Model*.....72
4.9. Perbandingan Hasil Klasifikasi Antara Pelabelan Data Training Manual dan Rulebase Pada Proses *Naïve Bayes Classifier* 75
4.9.1 Skenario Proporsi Pertama 75 % Training 25 % Testing 76
BAB 5 PENUTUP 78

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.1 Kesimpulan 78
5.2. Saran 79
Daftar Pustaka 80

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Daftar Gambar

Gambar 3.1 Metodologi Penelitian…….……29

Gambar 3.2. Proses Perhitungan Metode *Simple Weighted Product Model*……..…*40*

Gambar *4.1*. Contoh Data untuk Joko Widodo…….…..43

Gambar 4.2. Contoh Data Untuk KH.Makruf Amin…….……..44

Gambar 4.3 Contoh Data Untuk Prabowo Subianto……...45

Gambar 4.4. Contoh Data Untuk Sandiaga Salahudin Uno…….46

Gambar 4.5. *Source Code* Proses Prepocessing Data……..47

Gambar 4.6. Pembobotan Kata untuk Joko Widodo……..……...48

Gambar 4.7. Bobot Probabilitas Kata Sentimen Positif untuk Joko Widodo……..49

Gambar 4.8. *Woldcloud* sentimen positif Joko Widodo……..49

Gambar 4.9. Bobot Probabilitas Kata Sentimen Negatif untuk Joko Widodo……..50

Gambar 4.10. *Woldcloud* Sentimen Negatif Joko Widodo……50

Gambar 4.11 *Confusion Matrix* Proporsi Pertama Urut……..52

Gambar 4.12. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*…....53

Gambar 4.13. *Confusion Matrix* Proporsi Pertama Acak……...……54

Gambar 4.14. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*…...55

Gambar 4.15. *Confusion Matrix* Proporsi Kedua Urut……58

Gambar 4.16. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*……*59*

Gambar 4.17. *Confusion Matrix* Proporsi Kedua Acak……..59

Gambar 4.18. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*……*61*

Gambar 4.19. *Confusion Matrix* Proporsi Ketiga Urut……...63

Gambar 4.20. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*……*64*

Gambar 4.21. *Confusion Matrix* Proporsi Ketiga Acak……64

Gambar 4.22. Perbandingan Confusion Matrix Antara *Rulebase* dan *Naïve Bayes*…...66

Gambar 4.23. Tabel Jumlah Sentimen Positif dan Negatif ……..……..69

Gambar 4.24. Diagram Rangking Tokoh Publik……...……..…….....74

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Daftar Tabel

Tabel 2.1. PenelitianTerdahulu……7

Tabel 2.2 Kombinasi Awalan dan Akhiran yang Dilarang……..21

Tabel 2.3 Tabel *Confussion Matrix*……..23

Tabel 3.1 Contoh Data Uji Twitter……..31

Tabel 3.2 Teks Twitter hasil *Case Folding*……..33

Tabel 3.3 Term Hasil Proses Tokeninzing……..34

Tabel 3.4 Term Hasil Proses *Filtering*……35

Tabel 3.5 Term Hasil Proses *Stemming*……36

Tabel 3.6 Bobot Probabilitas *Naive Bayes* Untuk Sentimen……...38

Tabel 4.1. *Confusion Matrix* Proporsi Pertama Urut……...53

Tabel 4.2. *Confusion Matrix* Proporsi Pertama Acak……..54

Tabel 4.3.*Confusion Matrix* Proporsi Kedua Urut……..58

Tabel 4.4. *Confusion Matrix* Proporsi Kedua Acak……60

Tabel 4.5. *Confusion Matrix* Proporsi Ketiga Urut……63

Tabel 4.6. *Confusion Matrix* Proporsi Ketiga Acak……64

Tabel 4.7 Presentasi Nilai Akurasi Pada Proporsi Pertama……67

Tabel 4.8 Presentasi Nilai Akurasi Pada Proporsi Kedua……..68

Tabel 4.9 Presentasi Nilai Akurasi Pada Proporsi Ketiga……..69

Tabel 4.10 Tabel Alternatif Perhitungan SWP……..72

Tabel 4.8. Tabel Kriteria Perhitungan SWP……72

Tabel 4.9. Nilai Kriteria Setiap Alternatif Perhitungan SWP……..73

Tabel 4.13. Proporsi Pertama Urut Pada Pelabelan Secara Manual……76

Tabel 4.14. Proporsi Pertama Acak Pada Pelabelan Secara Manual……...76

Tabel 4.15. Proporsi Pertama Urut Pada Pelabelan Secara *Rulebase*……..76

Tabel 4.16. Proporsi Pertama Acak Pada Pelabelan Secara Rulebase……77

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Daftar Lampiran

Lampiran 1 Pembobotan Kata untuk KH. Ma'ruf Amin 82
Lampiran 2 Pembobotan Kata untuk Prabowo Subianto.....85
Lampiran 3 Pembobotan Kata untuk Sandiaga Salahudin Uno.....88
Lampiran 4 Skenario Proporsi Pertama 75 % Training 25 % Testing 91
Lampiran 5 Skenario Proporsi Kedua 80 % Training 20 % Testing 101
Lampiran 6 Skenario Proporsi Ketiga 90 % Training 10 % Testing 111
Lampiran 7 Contoh Sentiword Positif Dan Negatif…….....122
Lampiran 8 Sourc Code ……..….....126
Lampiran 9 Contoh Pengambilan Data Secara Urut dan Acak…….....130
Lampiran 10 Hasil Perbandingan Antara Rolebase dan Naïve Bayes….….131

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah suatu negara dengan pengguna internet terbesar di dunia. Hasil survey APJII (Asosiasi Penyelengara Jasa Internet Indonesia) pada tahun 2018, Kosumen internet di Indonesia meningkat sebanyak 8% menjadi 143,26 juta jiwa. Jumlah ini sama dengan 54,68% dari populasi 260 juta orang. Tingginya penggunaan internet di Indonesia disebabkan oleh berbagai alasan yang sangat beragam, menurut hasil survey APJII (Asosiasi Penyelengara Jasa Internet Indonesia) pada tahun 2018 pembaruan informasi menjadi alasan yang paling tinggi dalam mengakses internet. Untuk mendapatkan informasi mengenai pembaruan ini, pengguna internet di Indonesia mengakses berbagai konten yang tersedia. Ternyata hasil dari survey membuktikan bahwasahnya layanan yang paling sering diakses oleh pengguna adalah aplikasi chatting dengan 89,35%. Disusul dengan media sosial sebanyak 87,13%, mesin pencari (Google,dll) 74,84%, melihat gambar atau foto 72,79%, melihat video 69,64%, dan sisanya adalah aktivitas internet lain (APJII, 2019). Hal ini menunjukkan bahwa adanya tingkat penggunaan media sosial untuk saat ini sangat berkembang pesat seiring dengan jumlah pengguna yang terus menerus meningkat.

Saat ini media sosial banyak diminati oleh berbagai kalangan, yang memungkinkan penggunanya untuk mengirimkan pesan secara bebas. Maka dari itu media sosial mempunyai peran yang penting dalam memberikan informasi yang cepat. Informasi yang dihasilkan dari media sosial juga sangat beragam seperti pendapat, komentar, kritik, baik yang bersifat positif maupun negatif. Banyaknya jumlah pengguna media sosial di Indonesia tentu akan memunculkan berbagai macam bentuk status maupun opini terhadap topik tertentu tanpa adanya batasan waktu dalam berbagai informasi pada media sosial. Di Indonesia twitter

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menjadi salah satu situs *microblogging* yang sangat digemari oleh masyarakat. Terbukti dari banyaknya penggunaan twitter yang hanpir mencapai 19,5 juta pengguna dari total 330 juta pengguna di dunia. Dengan memiliki tampilan yang *friendly*, yang terbatas hanya berisi 140 karakter. Masyarakat sudah terbiasa dalam menyampaikan pendapat atau opininya dalam media sosial termasuk dalam twitter, hal ini digunakan dalam menanggapi suatu kejadian atau peristiwa yang terjadi di lingkungan mereka. Hal ini juga menjadi salah satu contoh untuk mengetahui sejauh mana sentimen masyarakat terhadap lingkungan mereka baik berupa kritik ataupun saran. (Ariadi & Fithriasari, 2015).

Analisis sentimen berguna dalam pemantauan media sosial, karena memungkinkan untuk mendapat *insight* dan gambaran umum dari opini publik yang lebih luas dibalik sebuah topik tertentu. Sentimen cukup mudah untuk dimengerti. Itu merupakan wujud perasaan atau emosi, sikap dan opini. Di media sosial, sentimen sebuah postingan bisa dilihat dengan nada atau emosi yang disampaikan terhadap sebuah *brand*. Namun, analisis sentimen pada umumnya memiliki beberapa tantangan, antara lain, keberagaman bahasa yang digunakan oleh masyarakat Indonesia untuk menyampaikan *tweet* mereka, yang mengakibatkan masalah sekaligus tantangan terhadap proses klasifikasi tulisan maupun opini yang ada di media sosial yang menggunakan Bahasa Indonesia. Analisa sentimen jika digunakan tepat pada sasaran, maka dapat digunakan untuk bisnis dan lebih proaktif dalam melihat perubahan dinamika pasar dan pelanggan. Maka dari itu, saat ini dibutuhkan suatu alat yang dapat menunjang dan mempermudah proses analisis sentimen, sehingga analisisnya dapat dilakukan dapat lebih cepat dan efisien.

Tahun 2019 merupakan tahun politik, dimana dilaksanakan pemilihan presiden dan wakil presiden Indonesia serta anggota legislatif. Dalam pelaksanaan pemilihan presiden dan wakil presiden ada berbagai macam opini dari masyarakat baik berupa sentimen positif maupun sentimen negatif untuk para tokoh publik yang mencalonkan diri untuk menjadi Presiden dan Wakil

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Presiden Republik Indonesia. Adapun tokoh publik tersebut adalah Bapak Joko Widodo, KH Ma'ruf Amin, Prabowo Subianto, dan Sandiaga Salahudin Uno. Pada penelitian kali ini peneliti akan meneliti sejauh mana citra positif yang dimiliki oleh keempat tokoh publik tersebut berdasarkan sentimen dari masarakat. Pada penelitian ini sentimen terbagi menjadi dua kelas yaitu sentimen negatif dan sentimen positif.

Permasalahan yang terjadi saat ini adalah bagaimana cara untuk menentukan kelas sentimen dari *tweet* terhadap beberapa tokoh. Klasifikasi dan sentimen terhadap tokoh publik tersebut menjadi perlu untuk diketahui karena dapat digunakan sebagai bahan analisa serta evaluasi bagi tokoh publik yang bersangkutan untuk mengukur citra positif dari tokoh tersebut. Pada saat akan melakukan proses analisis dan mengklasifikasi sentimen, tidak jarang masalah akan muncul jika proses tersebut dilakukan secara manual. Dengan adanya masalah tersebut dapat dipastikan akan membutuhkan waktu yang cukup lama untuk melakukan proses klasifikasi dokumen twitter tersebut, dan dengan menggunakan cara manual tentu tidak akan efektif. Untuk mengatasi masalah tersebut, solusi yang dapat digunakan yaitu berupa pembuatan sistem analisis sentimen yang bersifat positif dan negatif berkaitan dengan opini masyarakat terkait sejauh mana citra positif dari para tokoh publik yang bersangkutan.

Penelitian yang akan dilakukan adalah pengklasifikasian topik sentimen dan sentimen dari informasi teks twitter yang membahas mengenai tokoh publik menggunakan metode *Naïve Bayes Classifier dan Simple Weighted Product Model*. *Metode Naïve Bayes Clasifier* juga digunakan dalam menentukan sentimen dari *tweet* sehingga dapat diketahui sentimen positif, negatif yang menjadi kelas sentimen pada teks twitter tersebut. Sedangkan *Simple Weighted Product Model* digunakan dalam menentukan rangking atau urutan dari para tokoh publik yang bersangkutan. Konsep dasar metode *Simple Weighted Product Model* menggunakan perkalian untuk menghubungkan rating atribut, di mana rating setiap atribut harus dipangkatkan dengan bobot atribut yang bersangkutan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

API (*Application Programming Interface*) dari twitter digunakan untuk pengambilan data. Pemberian label kelas dengan jumlah yang besar dinilai membuat masalah yang cukup sulit. Untuk mengatasi masalah tersebut digunakan suatu metode untuk memudahkan dalam melakukan proses klasifikasi dengan jumlah data yang banyak. Sentiword posititif dan sentiword negative Bahasa Indonesia yang didapatkan dari github digunakan sebagai media dalam pelabelan otomatis berdasarkan penanda-penanda tertentu tersebut. Metode klasifikasi yang digunakan pada penelitian ini adalah metode klasifikasi *Naive Bayes Classifier*. Metode *Naive Bayes Classifier* merupakan salah satu dari metode klasifikasi yang mempunyai performa yang cukup baik untuk mengklasifikasikan data teks. Digunakan *Simple Wighted Product Model* sebagai suatu metode untuk memperoleh nilai rangking terbaik dari keempat tokoh tersebut.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi dari *Naïve Bayes Classifier* guna mengklasifikasi sentimen dari twitter pada opini masyarakat terkait dengan citra positif tokoh publik?
2. Bagaimana hasil peringkat atau urutan citra positif tokoh publik berdasarkan kelas sentimen masyarakat menggunakan metode *Simple Wighted Product Model*?
3. Bagaimana perbandingan hasil klasifikasi antara pelabelan data training yang dilakukan secara manual dan menggunakan *rulebase* pada proses *Naïve Bayes Classifier* ?

1.3 Batasan Masalah

Batasan masalah dalam penelitian kali ni yaitu:

1. Data yang digunakan yaitu opini masyarakat yang didapatkan dari media sosial twitter pada periode Januari – Mei 2019.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Penelitian ini sebatas hanya melakukan analisis sentimen terhadap *tweet* yang berbahasa Indonesia.
3. Tokoh publik yang dinilai dalam penelitian ini hanya dibatasi 4 tokoh yaitu Bapak Joko Widodo, KH Ma'ruf Amin, Prabowo Subianto, dan Sandiaga Salahudin Uno.
4. Penelitian ini tidak mencangkup kata dan kalimat yang cara penulisannya tidak umum (contoh: singkat, alay, emotikon).
5. Output klasifikasi dibagi kedalam dua kategori sentimen yaitu sentimen negatif dan sentimen positif.
6. Hasil dari penelitian ini hanya menggambarkan presepsi masyarakat secara umum. Tidak menggambarkan presepsi masyarakat secara khusus.
7. Tidak menanangani Sarkasme.
8. Untuk tweet yang membahas keempat tokoh, maka tweet tersebut akan masuk ke dalam setiap data set dari masing masing tokoh yang bersangkutan.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengetahui sejauh mana penerapan dari metode *Naive Bayes Classifier* untuk mengklasifikasikan sentimen pada opini masyarakat terkait dengan citra positif tokoh publik.
2. Tujuan dari penulisan skripsi ini adalah memberikan rangking (urutan) citra positif pada tokoh publik berdasarkan sentimen yang didapatkan menggunakan metode *Simple Weighted Product Model*.
3. Mengetahui sejauh mana perbandingan hasil nilai akurasi yang dihasilkan melalui pelabelan data training secara manual dan pelabelan data *training* menggunakan *rulebase* menggunakan *Naïve Bayes Classifier*

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. Untuk menambah wawasan serta pengetahuan mengenai implementasi *Naïve Bayes Classifier* dan *Simple Weighted Product Model.*
2. Sebagai salah satu bahan rujukan dalam penelitian mengenai klasifikasi teks.
3. Sebagai bahan rekomendasi kepada masyarakat maupun pihak pihak terkait terhadap tokoh publik yang memiliki citra positif terbaik berdasarkan informasi dari mesia sosial twitter.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 2

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Mempelajari penelitian dari pendahulu juga dapat membantu peneliti untuk menentukan bagaimana cara pengolahan data maupun analisis data, yaitu berdasarkan perbandingan terhadap apa saja yang telah dilakukan para peneliti sebelumnya, pada Tabel 2.1 Terdapat penelitian terdahulu yang sudah pernah dilakukan.

Tabel 2.1. Penelitian Terdahulu

No	Judul	Kesimpulan	Pembeda
1.	Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naïve Bayes Classifier di Forum Kaskus. Oleh Faishal Nuruz Zuhri/ 2017 (Zuhri & Alamsyah, 2017)	Penelitian tersebut telah berhasil mengklasifikasikan 6338 data uji berupa komentar di kaskus kedalam dua kelas sentimen yakni sentimen positif dan sentimen negatif. Data uji yang masih berupa teks normal dilakukan *prepocessing* menjadi potongan kata (*term*). Hasil dari *prepocessing* digunakan untuk menghitung bobot kemungkinan dari semua kelas sehingga bobot yang tertinggi akan menjadi kelas terpilih dari teks tersebut. Penelitian ini menggunakan metode *Naïve Bayes Classifier*	Pada penelitian ini, peneliti meneliti sejauh mana sentimen masyarakat terkait dengan brand Smartfren yang beredar di pasaran. Peneliti menggali data pada aplikasi Kaskus.
2.	Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter Oleh : Yislam (2016) (Ui, 2016)	Berdasarkan hasil klasifikasi sentimen, dibidang Politik dan Ekonomi sentimen positif lebih dominan dibandingkan sentimen negatif. Akan tetapi, berbeda hal dengan Hukum yang mendapatkan sentimen negatif. Hal ini mungkin terjadi karena adanya berbagai permasalahan korupsi di Indonesia. Pada bidang Politik, tweet yang sering ditulis mengenai penolakan revisi UU KPK oleh DPR. Terakhir, pada bidang Ekonomi, permasalahan yang muncul pada	Pada penelitian tersebut, peneliti meneliti sejauh mana sentimen masyarakat terhadap masa pemerintahan Jokowi menggunakan data twitter dengan mengacu pada beberapa aspek yaitu politik,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		tweet yakni masalah krus rupiah ke dollar, kenaikan harga BBM. Hasil klasifikasi sentimen yang didapatkan sebesar 51.29% bersentimen netral. Hal ini desebabkan oleh beberapa factor yaitu masih terdapatnya duplikasi pada *tweet* dan *tweet* berasal dari *retweet* ataupun hasil dari *shere* berita online ke *tweet.*	ekonomi, dan hukum.
3.	Analisis Sentimen Berdasarkan Komentar Publik Terhadap Toko Online Zalora dan Bery Benka pada Media Sosial Facebook. Oleh Gusriani, dkk/ 2016 (Syahmia Gusriani, Kartina Diah Kusuma Wardhani, 2016)	Proses pertama untuk melakukan analisis sentimen dimulai dari data *collecting, preprocessing, feature selection,* klasifikasi dan pengukuran akurasi. Hasil dari analisis pengujian menunjukkan bahwa *Naïve Bayes Classifier,* memiliki tingkat kestabilan akurasi setelah melalui pengujian dengan beberapa nilai *Frequent Itemset*. *Naïve Bayes Classifier* memiliki rata rata akurasi 90,3%. Penelitian ini menggunakan metode *Naïve Bayes Classifier, K-NN*, dan *Decision Tree.*	Pada penelitian ini, peneliti meneliti sejauh mana komentar publik terkait toko online yang beredar di pasaran. Peneliti menggali data pada media sosial facebook. Penelitian ini menggunakan metode *K-NN*, dan *Decision Tree.*
4.	Analisis Sentimen Masyarakat Terhadap e-Commerce Pada Media Sosial Menggunakan *Metode Naïve Bayes Classifier* (NBC) Dengan Seleksi Fitur *Information Gain* (IG). Oleh Abdan Syakuro/2017 (SYAKURO, 2017)	*Naïve Bayes Classifier* dengan menggunakan seleksi fitur *Information Gain* sudah terbukti untuk menganalisis sentimen secara otomatis. Percobaan yang dilakukan pada data testing secara realtime, setiap data diklasifikasikan menjadi dua yaitu sentimen positif dan sentmen negatif. Performa dari penggabungan dua metode *Naïve Bayes Classifier* dengan seleksi *fitur Information* gain mengalami suatu kenaikan yaitu pada batas threshold 2200 waktu proses menjadi 514 detik lebih cepat dari sebelum penambahan *Information Gain*. Jadi dapat disimpulkan bahwa suatu sistem analisis sentimen jika ditambahkan dengan seleksi fitur *Information Gain* dapat menambah akurasi hingga 88,8%.	Studi kasus yang diambil dalam penelitian ini adalah *E-Commerce* yang meliputi Lazada, Bukalapak, dan Tokopedia. Metode yang digunakan yaitu *Naïve Bayes Classifier* dengan menambahkan seleksi fitur *Information Gain*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.	Analisis Sentimen Komentar Pada Sistem Penilaiyan Kinerja Instruktur Training ICT (*Information and Communication Technologogy*) UIN Sunan Kalijaga Menggunakan *Naïve Bayes Classifier* Oleh Laksminta Sasti/ 2017	Berdasarkan penelitian yang sudah dilakukan, maka kesimpulan yang diperoleh adalah penggunaan metode *Naïve Bayes Classifier* telah berhasil melakukan analisis sentimen dan klasifikasi komentar yang berbahasa Indonesia pada sistem penilaiyan kinerja instruktur training ICT UIN Sunan Kalijaga dengan akurasi pengujian mencapai 85,31%. Penelitian ini menggunakan metode *Naïve Bayes Classifier*.	Penelitian ini meneliti analisis sentimen dari mahasiswa UIN Sunan Kalijaga melalui kolom komentar pada aplikasi ICT.
6.	Sistem Analisis Sentimen Publik Tentang Opini Pemilihan Kepala Daerah Jawa Timur 2018 Pada Dokumen Twitter Menggunakan *Naïve Bayes Classifier* Oleh Fajar Darwis Dzikril Hakimi/ 2018	Berdasarkan penelitian yang sudah dilakukan, metode naïve bayes lassifier merupakan metode yang dapat diterapkan untuk melakukan klasifikasi terkait dengan opini pemlihan Kepala Daerah Jawa Timur 2018. Bisa dibuktikan dengan tingginya peforma sistem yang dihasilkan yaitu menghasilkan akurasi sebesar 98,99%, presisi sebesar 93,44%, recall sebesar 97,78%, dan f-measure sebesar 95,56%. Dan kecendrungan masyarakat yaitu lebih cenderung memilih calon gubernur nomer urut satu dibandingkan calon gubernur nomor urut dua.	Penelitian ini menggunakan klasifikasi secara manual dan tidak menggunakan rolebase sebagai bahan acuan dalam melakuakan klasifikasi. Studi kasus yang digunakan yaitu tentang opini masarakat terkait dengan pemilihan Kepala Daerah Jawa Timur 2018.

2.2 Landasan Teori

Bab ini membahas mengenai penelitian terdahulu, teori yang berkaitandengan penelitian, dan metodologi yang relevan dengan penelitian ini. Pada bagian akhir, seluruh yang telah dibahas sebelumnya dirangkum menjadi suatu kerangka teoritis yang dijadikan dasar keseluruhan penelitian.

2.2.1 *Machine Learning*

Algoritma *Machine Learning* adalah suatu algoritma yang mempunyai kemampuan untuk belajar dari data dan pengalaman dengan memperhatikan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

beberapa hal, contohnya seperti tugas dan ukuran kinerja, jika suatu tugas dapat diukur untuk meningkatkan pengalaman, sehingga mesin tersebut dapat dikatakan belajar (Mitchell, 2009).

Tantangan utama dalam *Machine Learning* adalah algoritma harus dapat berkerja dengan baik pada hal hal yang sebelumnya belum pernah diketahui, tidak hanya pada model yang digunakan. Kemampuan untuk berkinerja baik pada input yang sebelumnya tidak diketahui disebut dengan generalisasi. Namun, masalah yang sering kali terjadi dalam *Machine Learning* adalah bahwa set pelatihan yang besar diperlakukan untuk generalisasi yang baik, namun rangkaiyan pelatihan yang besar juga lebih mahal secara komputasi. Segingga terdapat faktor-faktor yang dapat digunakan untuk menentukan seberapa baik kemampuan suatu algoritma *Machine Learning* yaitu:

1. Membuat nilai kesalahan yang dihasilkan data pelatihan kecil.
2. Membuat jarak antara nilai kesalahan yang digasilkan oleh data pelatihan dan data test kecil.

Terdapat dua faktor yang sesuai dengan tantangan di dalam *Machine Learning* ini yaitu *underfitting* dan *overfitting* (Goodfellow, Bengio, & Courville, 2016). *Underfitting* terjadi ketika suatu model tidak dapat memperoleh nilai kesalahan yang rendah pada set pelatihan. Sedangkan, *overfitting* terjadi jika jarak antara nilai kesalahan yang dihasilkan data pelatihan dan uji terlalu besar.

2.2.2. Analisis Sentimen

Analisis Sentimen mengacu pada bidang yang sangat luas mulai dari penerapan bahasa alami, komputasi *linguistic* dan *text mining* yang bertujuan untuk menganalisa sikap, evaluasi, sentimen, pendapat penilaiyan serta emosi seseorang yang berkaitan dengan suatu produk, layanan, topik, individu, maupun kegiatan tertentu (Chen, Su, & Chen, 2011).

Analisis sentimen memiliki tugas utama yaitu untuk mengelompokkan text yang sudah ada pada sebuah dokumen atau kalimat, lalu menentukan pendapat yang terkandung dalam kalimat atau dokumen tersebut, apakah termasuk ke

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dalam sentimen positif, negatif, atau netral. Analisis sentimen juga dapat menyatakan perasaan emosional seperti gembira,sedih, ataupun marah.

Analisis sentimen memiliki tujuan yaitu untuk menentukan opini atau perilaku dari seorang dengan memperhatikan suatu topik tertentu. Tergantung dari (bagaimana si penulis ingin mempengaruhi pembaca).

2.2.3. *Text Mining*

Text Mining berhubungan dengan penciptaan informasi baru secara otomatis atau semi otomatis yang sebelumnya tidak diketahui dari data berkualitas tinggi dari sejumlah besar Text yang tidak terstruktur (Hearst & Hall, n.d.). Dalam menjalankan proses *Text Mining* terdapat tiga teknik yang berurutan, masing masing dapat dikaitan dengan *Text Mining*. Teknik tersebut merupakan penggabungan dari information retrieval, *natural language processing*, dan data mining. Pertama *information retrieval* bertugas dalam mengumpulkan text yang berpotensi dan relevan terhadap tugas yang diberikan. Kedua *Natural Language Processing* bertugas dalam menganalisis teks masukan untuk mengidentifikasi dan menyusun informasi yang relevan. Ketiga Data Mining bertugas dalam menemukan pola dalam informasi terstuktur yang telah disimpulkan dari teks.

Aspek utama *Text Mining* sebenarnya sama dengan yang dipelajari dalam *empirical computational linguistic*. Meskipun focus pada *Natural Languange Processin*g, beberapa masalah yang terkait dengan lingistik komputasi juga dibahas dalam i*nformation retrieval* dan data mining.

a. *Information Retrieval*

Information Retrieval merupakan suatu metode untuk mencari dan mendapatkan teks teks dari sekumpulan besar teks yang tidak terstruktur yang dapat memenuhi kebutuhan informasi, berdasarkan *query* yang diberikan (Manning, Raghavan, & Schutze, 2008). Ada beberapa konsep dalam Information *Retrieval* yang berkaitan dengan analisis teks antara lain:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. *Vectors*, Untuk menentukan relevansi teks, banyak pendekatan memetakan semua teks dan kueri ke dalam model ruang *vector* (Manning et al., 2008). Model seperti itu mendefinisiakan representasi *vector* umum x = (x_1,.....,x_k), $k \geq 1$, untuk semua masukan dimana masing-masing $x_i \in x$ mempresentasikan *property* masukan. Masukan konkret seperti text D kemudian ditunjukkan oleh satu nilai $x_i^{(D)}$ untuk masing masing x_i.
2. *Similarity* yaitu dengan memberikan representasi umum, kesamaan antara teks dan query yang dapat dihitung. Dengan adanya representasi umum, kesamaan antara teks dan *query* dapat dihitung. Sebagian besar frekuensi kata dari *query* pencarian akan sering menghasilkan nilai 0.
3. *Indexing* yaitu semua teks dalam koleksi tertentu telah diindeks sebelumnya. Permintaan kemudian dicocokkan dengan indeks pencarian, sehingga menghindari memproses teks sebenarnya selama pencarian. Pendekatan pengindeksan yang sangat canggih ada dan digunakan di mesin pencari web saat ini (Manning et al., 2008). Dalam bentuk dasarnya, indeks pencarian berisi satu entri untuk setiap *property* terukur. Setiap entri menunjuk ke sumua teks yang relevan sehubungan dengan *property*.
4. *Filtering* dibagi mejadi dua tingkat *Text Filtering* dan *Passage Retrival*. *Teks Filtering* digunakan untuk mengklasifikasi teks lengkap sebangai data teks yang relevan digunakan atau tidak relevan untuk digunakan, sedangkan *Passage Retrieval* bertujuan untuk menentukan bagian teks yang relevan untuk menjawab query tertentu.

b. *Natural Languange Processing*

NLP dapat artikan sebagai sebuah proses bahasa manusia secara otomatis atau semi otomatis (Copestake, 2007). Proses tersebut dilakukan untuk mengkaji sejauh mana interaksi computer dengan menggunakan Bahasa alami manusia yang digunakan dalam berinteraksi sehari hari.

Pada proses *natural language processing* ada beberapa kesulitan yang sering terjadi diantaranya adalah terjadinya ambiguitas yang memungkinkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

adanya intrepetasi yang berbeda, sehingga algoritma analisis teks perlu untuk dapat menyelesaikan masalah ambiguitas. Bagaimanapun, seringkali perkataan manusia tidak selalu tepat, seringkali ambigu dan struktur tata Bahasa dapat bergantung pada banyak variable yang kompleks, termasuk Bahasa gaul, dialek daerah, dan konteks sosial.

c. Data *Mining*

Data *mining* dapat juga disebut sebagai penemuan pengetahuan dalam database. Hal ini biasanya didefinisikan sebagai proses untuk mendapatkan pola atau ilmu yang berguna dari sumber data (Chen et al., 2011). Pola harus valid, berpotensi bermanfaat, dan mudah dimengerti. Data mining bertujuan untuk mengumpulkan informasi baru dari jenis informasi yang telah ditentukan dari sejumlah besar data masukan yang biasanya besar, yang sudah diberikan dalam bentuk terstruktur (Witten, Frank, & Hall, 2011). *Machine Learning Algoritma* mengenali pola statistik dalam kasus yang relevan untuk masalah prediksi, proses ini disebut *training*. Pola yang ditemukan kemudian digeneralisasi sehingga bisa diterapkan untuk menyimpulkan informasi baru dari data yang tidak terlihat, yang umumnya disebut sebagai prediksi. Dalam hal ini, *Machine Learning* dapat dilihat sebagai dasar teknis aplikasi data mining (Witten et al., 2011).

Data Mining dan *Text Mining* terkait memiliki ketertarikan dalam dua hal. Pertama, informasi keluaran tersruktur dari analisis teks berfungsi sebagai masukan untuk *Machine Learning*. Untuk melatih pengklasifikasian text. Kedua, banyak analisis text itu sendiri bergantung pada algoritma Machine Lerning untuk menghasilkan informasi keluaran.

2.2.4 *Naïve Bayes Classifier*

Metode *Naïve Bayes Classifier* adalah metode yang dipakai sebagai cara untuk memperkirakan suatu probabilitas (Rini, Farida, & Puspitasari, 2016). *Naive Bayes Classifier* merupakan sebuah konsep dari probabilitas yang dipakai sebagai cara untuk menentukan kelompok kelas dokumen teks dan juga dapat

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

juga mengolah data dalam jumlah besar serta menghasilkan akurasi yang cukup tinggi (Lestari, Perdana, & Fauzi, 2017). Tingkat peforma dari sistam klasifikasi yang dibuat menggunakam *Naïve Bayes Classifier* tergantung pada data yang dimiliki dan data yang diambil sebagai data *training.* Jika data yang diambil sebagai data *training* bisa mewaliki semua ataupun sebagian besar data yang dimilikii, maka sistem klasifikasi mempunyai peforma yang baik. Ketika suatu sistem klasifikasi yang dibuat mempuyai peforma yang baik, maka sistem tersebut bisa diterapkan untuk melakukam klasifikasi terhadap data yang lebih banyak.

Tahapan proses *Naïve Bayes Clasiffier* adalah:

1. Menentukan data uji.
2. Menentukan probabilitas kategori dari data latih.

   $$p(ci) = \frac{jumlah\ dokumen\ dengan\ label\ ci}{keseluruhan\ dokumen} \quad (3)$$

3. Menghitung probabilitas kosakata (*term frequency*) dari data latih.
4. Menghitung probabilitas untuk setiap kelas data latih

   $$p(wkj.ci) = \frac{pk+1}{n+term} \quad (4)$$

   Dimana,

   p(wkj.ci) = probabilitas term

   pk + 1 = fitur kemunculan term wkj +1, penambahan 1 dilakukan agar hasil perhitungan tidak 0 (nol).

   n = semua jumlah kata yang ada pada topik ci

   term = jumlah kata keseluruhan.
5. Menentukan Vmap untuk data uji.

   $$Vmap = p(ci) \times p(wkj.ci) \quad (5)$$

6. Klasifikasi Data Uji

   Jadi dari perhitungan diatas data uji tersebut termasuk dalam kaegori kelas negatif (-) karena Vmap Negatif lebih besar (>) dari Vmap Positif.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.5. *Simple Weighted Product Model*

Simple Weighted Product Model adalah suatu metode penyelesaian memakai perkalian yang digunakan untuk menghubungkan rating dari masing masing atribut. Dimana rating tersebut harus dipangkatkan terlebih dahulu menggunakan bobot atribut yang ada. Proses ini disebut dengan proses normalisasi. *Simple Weighted Product Model* adalah keputusan analisis multi-kriteria yang popular dan merupakan metode pengambilan keputusan multi-kriteria. Seperti semua metode FMADM untuk menyelesaikan kasus kasus dimana data terdiri atas banyak atribut kepentingan terdiri dari metode *Simple Additive Weighting Method* (SAW), *Simple Weighted Product Model* (SWP), ELECTRE, TOPSIS, dan Analytic Hierarchy Proses (AHP).

Pemilihan metode *Simple Weighted Product Model* didasarkan juga atas kemampuannya dalam memberikan solusi optimal dalam system pemeringkatan. Pemilihan metode ini juga didasarkan atas kompleksitas komputasi yang tidak terlalu sulit sehingga waktu yang dibutuhkan dalam menghasilkan perhitungan relative singkat.

Metode *Simple Weighted Product Model* juga telah banyak digunakan sebagai referensi dalam sistem pemeringkatan dan Sistem Penunjang Keputusan (SPK), sebagaimana dilakukan oleh Ahmadi, A., dan Wiyanti, DT., tahun 2014 untuk menentukan daftar penerimaan bantuan Langsung Masyarakat berdasarkan kriteria keputusan, juga oleh Nurjannah, N., Arifin, Z., Khairina, DM., tahun 2015 yang menggunakan metode *Simple Weighted Product Model* untuk SPK pembelian sepeda motor.

Preferensi untuk alternatif diberikan sebagai berikut:

$$S_i = \prod_{J=1}^{n} X_{ij}^{\,w_j} \quad ; \text{dengan} \quad i = 1, 2, \ldots\ldots, \quad m \tag{6}$$

$$W_J = \frac{w_j}{\sum w_j} \tag{7}$$

$$\sum W_j = 1 \tag{8}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Preferensi relative dari setiap alternatif, diberikan sebagai

$$V_i = \frac{\prod_{j=1}^{n} x_{ij}\ wj}{\prod_{j=1}^{n} (X_j *)wj} \quad \text{dengan} \quad I = 1, 2, \ldots\ldots, \quad m \qquad (9)$$

Dimana:

S : Preferensi alternatif dianalogikan sebagai vektor S

x : Nilai kriteria.

w : Bobot kriteria.

i : Alternatif.

j : Kriteria.

n : Banyaknya kriteria.

m : Banyaknya alternatif.

W_j : Pangkat yang mempunyai nilai positif untuk atributkeuntungan, dan memiliki nilai negatif untuk atribut biaya.

V : Menyatakan preferensi alternatif dianalogikan sebagai vector V.

Langkah – langkah Metode Simple Weighted Product Model

Secara singkat algoritma Simple Weighted Product Model adalah:

1. Memberi nilai dari setiap alternatif (Ai) pada setiap kriteria (Cj) yang ditentukan, di mana nilai i=1,2,……m.
2. Melakukan perbaikan pembobotan untuk menghasilkan nilai $\Sigma W_j = 1$ dimana j = 1,2,……n.
3. Memberikan kategori untuk masing – masing kriteria, apakah masuk ke dalam kriteria keuntungan atau kriteria biaya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4. Memberikan nilai vektor S dengan cara mengalikan semua kriteria pada sebuah alternatif dengan bobot sebagai pangkat positif untuk kriteria keuntungan dan bobot sebagai pangkat negatif pada kriteria biaya.
5. Memberikan nilai vektor V yang berguna pada proses perangkingan.
6. Membandingkan nilai akhir dari vektor V.
7. Menemukan nilai urutan dari alternatif terbaik yang akan menjadi sebuah keputusan.

2.2.6. Twitter

Twitter menjadi salah satu media sosial yang memungkinkan penggunanya untuk mengirim serta membaca pesan *tweet* berupa teks, gambar atau sebuah video. Media sosial twitter berbeda dengan media soaial lainnya terutama dalam penulisan status atau *tweet*. Media sosial selain twiteer tidak ada batasan karakter yang dapat dituliskan sedangkan twitter hanya memberikan 280 karakter yang dapat ditulis sebagai status atau cuitan.

Twitter bersifat publik, maksudnya adalah semua yang dituliskan atau dibagikan dapat dilihat oleh semua pengguna lainnya, namun pengguna twitter dapat membatasi pengiriman *tweet* hanya dilihat temannya saja atau biasa disebut *follower*. Twitter memiliki fitur utama yakni dapat menuliskan status atu cuitan serta dapat melakukan pengiriman pesan kepada pengguna lain, fitur lain dari sosial media twitter sebagai berikut:

1. *Following:* Fitur untuk saling terhubung dengan pengguna lain atau bisa disebut sebagai pertemanan.
2. *Retweet:* Fitur *retweet* merupakan fitur yang memudahkan pengguna untuk meneruskan atau menyebarkan *tweet* pengguna lain
3. *Hastag: Hastag* atau tagar merupakan fitur dari twitter yang dapat mngelompokkan sebuah *tweet*.
4. *Trending Topic:* Fitur *trending* merupakan fitur yang menampilkan topik atau hastah yang sedang popular atau banyak dibahas oleh pengguna *tweet*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.7. Kamus Sentiword

Sentiword atau kata opini merupakan kumpulan kamus kata positif dan negatif dalam bahasa indonesia. Sentiword dibuat pertama kali oleh Liu dalam penelitiannya tetapi kamus *sentiword* yang tersedia adalah dalam Bahasa Inggris. Kamus *sentiword* yang digunakan dalam penelitian kali ini berasal dari Github yang sudah tersedia dalam bahasa Indonesia. *Sentiword* ini berguna untuk menghitung kata sentimen dalam kalimat.

2.2.8. Pre-Proses Teks

Pada tahapan pre-proses teks ini yaitu tahapan yang dilakukan untuk menyiapkan suatu dokumen teks agar siap untuk dilakukan proses klasifikasi. Tahapan ini dilakukan karena teks yang memiliki struktur yang tidak jelas tidak dapat di proses secara langsung. Alasan lain melakukan tahap ini adalah karena tidak semua kata dalam dokumen teks dapat mencermikan isi yang ada dalam dokumen tersebut. Tahapan ini dilakukam sebelum dokumen teks melakukan proses klasifikasi menggunakan metode *Naïve Bayes Classifier*. Tahapan pre-proses teks ini yaitu *cleansing*, *case folding*, *tokenisasi*, *filtering*, dan *stemming*.

1. *Cleansing*

   Tahapan pertama dalam pre-proses teks adalah *Cleansing*. Tahapan *Cleansing* ini digunakan untuk menghilangkan karakter web atau html yang ada tetapi tidak bermakna dalam proses klasifikasi dokumen. Hal ini terjadi karena dalam sebuah *tweet* tidak jarang masih terdapat suatu alamat html ataupun web yang jika tidak dihapus maka akan menganggu proses klasifikasi.

2. *Case Folding*

   Proses yang kedua dalam pre-proses teks adalah *Case Folding*. *Case Folding* merupakan suatu proses yang penting untuk dilakukan dalam melakukan klasifikasi dokumen. Cara melakukan proses *case folding* yaitu dengan mengubah semua huruf yang ada menjadi huruf kecil secara keseluruhan. Ini dilakukan bertujuan untuk menyamakan arti dari kalimat dari awal hingga

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

akhir (Manning et al., 2008). Selain itu, di dalam sosial media twitter. Ada beberapa masyarakat yang menuliskan *tweet* menggunakan huruf besar semua dengan maksud untuk melakuakan penekanan makna. Sehingga sulit jika data tersebut diklasifikasikan. Setelah proses *case folding* yaitu dengan menghapus semua tanda baca yang ada dan digantikan dengan karakter spasi.

3. Tokenisasi

   Proses yang ketiga dalam pre-proses teks adalah tokenisasi. Tokenisasi merupakan suatu proses untuk memilah suatu dokumen teks yang berupa kalimat ataupun pragraf dijadikan bagian-bagian tertentu (Manning et al., 2008). Tokenisasi perlu dilakukan untuk memudahkan pemberian bobot pada setiap katanya.. Contohnya adalah seperti pada kalimat " partai politik memiliki peran penting dalam penilaian masyarakat terhadap kreadibilitas tokoh publik yang diusungnya " setelah dilakukan tokenisasi maka menjadi " partai – politik – memiliki – peran – penting – dalam – penilaian – masyarakat – terhadap – kredibilitas – tokoh – publik – yang – diusungnya " yang terdiri dari 14 token.

4. *Filtering*

   Kalimat yang sudah melewati proses pemisahan kata melalui proses tokenezing selanjutnya akan dilakukan proses *filtering*. Proses akan menghapus kata kata yang tidak penting atau tidak mempunyai makna seperti kata sambung. Pada proses *filtering* dibutuhkan sebuah database yang menyimpan daftar *stopword* yang digunakan sebagai acuan dalam penghapusan kata. Kata hasil proses *tokenezing* digunakan sebagai inputan pada proses *filtering*. Setelah didapatkan kata tersebut maka tiap kata akan dilakukan pengecekan kata dengan pembanding yaitu kata yang ada dalam *stopiword*. Jika kata tersebut termasuk dalam kumpulan kata *stopword* maka dilakukan pengecekan kembali untuk jumlah kata. Jika dalam pengecekan ditemukan jumlah kata lebih dari satu maka kata tersebut akan dilakukan pengecekan *stopword* kembali.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5. *Stemming*

Proses setelah *filtering* adalah proses *stemming*. Proses *stemming* yaitu suatu proses pencarian kata dasar terhadap sebuah kata yang telah melalui proses *filtering*. Pada saat proses ini kata akan diubah menjadi suatu kata dasar sehingga kata yang memiliki imbuhan depan atau imbuhan belakang akan dihapus dan disesuaikan kata dasarnya. Proses *stemming* yang akan mengubah bentuk kata yang awalnya bukan kata dasar diubah menjadi kata dasar menggunakan metode Nazief Adriani.

1. *Nazief and Adriani's Stemmer*

Algoritma ini dikembangkan untuk mendapatkan kata dasar dengan menghilangkan imbuhan kata berdasarkan aturan pada Bahasa Indonesia yakni imbuhan awalan (prefiks), akhiran (sufiks), awalan dan akhiran (konfiks), dan sisipan (infiks). Algoritma ini juga dapat digunakan unruk *recording*, sebuah pendekatan untuk mengembalikan huruf awal kata yang terhapus akibat penghilangan prefix. Selain itu, algoritma ini juga menggunakan daftar kata dasar yang dipakai pada tahap pemeriksaan ketika proses *stemming* telah menemukan kata dasar yang diduga. Pengelompokan beberapa kategori aturan imbuhan dimodelkan pada algoritma *Nazief and Adriani's Stemmer* sebagai berikut:

$$\left[\left[\left[\ AW+\right]AW+\right]AW+\right]\ \text{Kata Dasar}\ \left[\left[+\ AK\right]\left[+KK\right]\left[+P\right]\right] \qquad (10)$$

Dimana:

AW = Awalan KK = Kata Ganti Kepunyaan

AK = Akhiran P = Partikel

Langkah langkah Nazief and Andriani's Stemmer adalah sebagai berikut (Asian,2007).

1. Kata yang belum melakukan proses stemming, dicari pada kamus kata dasar. Jika ditemukan, maka kata tersebut dianggap sebagai kata dasar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dan proses tersebut berhenti. Jika tidak ditemukan, maka dilanjutkan pada tahap selanjutnya.

2. Menghilangkan *inflection particle* (-lah, -kah, -tah, -pun) dan dilanjutkan menghilangkan *passive pronoun* (-ku, -mu, -nya).
3. Menghilangkan *derivation suffixes* (-I, -kan, -an).
4. Menghilangkan *derivation prefixes* (di-, ke-, se-, me-, be-, te-, pe-) dengan iterasi maksimum tiga kali dengan langkah langkah sebagai berikut.
   a. Langkah pertama akan berhenti jika:
      1. Terjadi suatu kombinasi awalan dan akhiran terlarang
      2. Awalan yang akan dideteksi saat ini sama dengan awalan yang dihilangkan sebelumnya.
      3. Tiga awalan telah dihilangkan.

Tabel 2.2 Kombinasi Awalan dan Akhiran yang Dilarang

Awalan	Akhiran yang Tidak Diperbolehkan
be-	-i
di-	-an
ke-	-I, -kan
me-	-an
se-	-i, -kan
te-	-ans

   b. Identifikasi tipe awalan kemudian hilangkan. Awalan terbagi menjadi dua tipe sebagai berikut:
      1. Standar (di-, ke-, se-) merupakan awalan yang dapat dihilangkan secara langsung dari kata.
      2. Kompleks (me-, be-, pe-, te-) merupakan awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya.
   c. Mencari kata apa saja yang telah dihilangkan awalannya dalam kamus kata dasar. Apabila tidak ditemukan, Maka keseluruhan tahapan dapat dihentikan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5. Jika kata dasar belum juga ditemukan, maka proses selanjutnya adalah *recording*. *Recording* dilakukan dengan menambah atau mengganti huruf awal kata yang terpenggal proses *stemming*. Contoh, kata ‘menangkis’ dimana selalu dihilangkan awalan ‘me-‘ menjadi ‘nangis’ tidak terdapat pada kamus kata dasar sehingga dilakukan recording dengan mengganti karakter ‘n’ menjad ‘t’ dan didapat kata dasar ‘tangkis’.

6. Jika semua cara atau tahapan gagal, maka input kata dianggap sebagai kata dasar.

6. Pembobotan Kata

Term (istilah) dapat berupa trase atau kata yang ada pada suatu dokumen, untuk bisa mengetahuii konten dari setiap dokumen tersebut.

TF (Terada diberi nilai 0m *Frequenccy*) merupakan suatu bentuk frekuensi dari kemunculan term pada dokumen. Jadi, pembobotan TF didasari pada jumlah kemunculan kata atau term pada sebuah kalimat.

2.2.9. Evaluasi Sistem Klasifikasi

Harapan dari dibuatnya sistem klasifikasi yaitu agar sistem bisa mengklasifikasikan data dengan baik dan benar (Prasetyo, 2014). Namun, pada kenyataannya, sebuah sistem pasti memiliki suatu kesalahan didalamnya. Maka langkah evaluasi sistem pada proses klasifikasi perlu dilakukan agar mengetahui seberapa bagus metode tersebut digunakan untuk melakukan proses klasifikasi.

Evaluasi sistem adalah suatu hal yang penting untuk dilakukan dalam setiap proses pembuatan sistem klasifikasi. *confusion matrix* merupakan suatu proses evaluasi sistem klasifikasi yang paling umum digunakan. *Confusion matrix* merupakan sebuah *matrix* yang dapat diterapkan untuk membandingkan hasil dari klasifikasi dengan data yang sebenarnya (Abdillah, 2011). *Confusion matrix* ini dapat digambarkan seperti pada Tabel 2.3

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 2.3 Tabel *Confussion Matrix*

Prediksi / Actual	Positif	Negatif
Positif	True Positif	False Negatif
Negatif	False Positif	True Negatif

Pada *confusion matrix* terdapat empat jenis data yaitu *true* positif, *true* negatif, *false* positif, dan *false* negatif (Lestari et al., 2017). *True* positif merupakan jumlah data yang memiliki kelas positif yang dikasifikasikan sebagai kelas positif. Sedangkan *true* negatif merupakan jumlah data yang memiliki kelas negatif yang diklasifikasikan sebagai kelas negatif. *True* positif dan *true* negatif adalah hasil klasifikasi yang benar. *False* positif adalah data yang sebenarnya yaitu kelas negatif tetapi diklasifikasikan sebagai kelas positif. Sedangkan *false* negatif adalah data sebenarnya yang merupakan kelas positif tetapi diklasifikasikan sebagai kelas negatif.

Akurasi, presisi,dan *recall* merupan tiga jenis evaluasi yang biasa digunakan untuk proses evaluasi sistem. Perhitungan tiga jenis evaluasi ini dilakukan untuk semua data uji yang tersedia. Akurasi sendiri yaitu perbandingan dari banyaknya data yang diklasifikasikan secara benar dibagi oleh jumlah data secara keseluruhan. Sedangkan presisi adalah perbandingan antara banyaknya *true* positif dibagi oleh jumlah semua data yang diklasifikasikan positif. *Recall* adalah perbandingan antara banyaknya *true* positif dibagi oleh jumlah semua data yang berlabel positif.

Akurasi menjadi perhitungan yang paling sering digunakan. Akurasi sendiri dapat dipakai untuk menghitung kinerja dari sistem klasifikasi jika data yang dipakai memiliki perbandingan jumlah kategori setiap data yang seimbang (Prasetyo, 2014) . Namum tidak jarang nilai akurasi tidak menggambarkan performa sebenarnya dari sebuah sistem klasifikasi. Hal ini bisa terjadi karena perbandingan dari jumlah kategori setiap data yang ada sangat tidak seimbang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

(Prasetyo, 2014). Oleh karena itu, menggunakan perhitungan akurasi saja dinilai masih belum cukup dikarenakan untuk melihat tingkat dari peforma sistem klasifikasi yang mempunyai proporsi kategori data yang seimbang. Sehingga, perlu dilakukannya perhitungan evaluasi lain agar bisa menggambarkan peforma sistem untuk setiap kategori yaitu dengan menggunakan *presisi* dan *recall*.

Untuk menghitung akurasi, presisi, dan *recall* dari sebuah sistem klasifikasi digunakan formula sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

Dimana :

TP = True Positif

TN = True Negatif

FP = False Positif

FN = False Negatif

2.2.10. *R Programing*

Software "R" adalah OSS untuk pengolahan data dan analisis statistic berbasis Bahasa program dan merupakan perangkat lunak yang menggunakan GUI (*Grafik User Interface*). "R" muncul setelah "S" dan "S plus" *software* berkembang. *Software* "R" dikembangkan sejak tahun 1995 OLEH Robert Gentleman dan Ross Ihaka dari Departemen Statistik Universitas Auckland, New Zealand. Sekarang, *software* ini sudah digunakan dan dikembangkan oleh banyak volunteer diberbagai penjuru dunia. *Software* "R" memiliki banyak fungsi, biasanya divisualisasikan dalam jemdela window atau disimpan dalam berbagai format seperti jpg, png, bmp, ps, emf, pictex dan xfif. *Softwere* "R" juga bias digunakan untuk ekstrak atau pengambilan data, msnsjemen data, penampilan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

data, pemodelan data, dan aplikasi data untuk berbagai anlisis. Hal penting yang terdapat pada R Programing adalah suatu lingkungan pemrograman yang bersifat mandiri yaitu dengan belajar sendiri (Borcard et al., 2011).

2.3 Integrasi Keilmuan

Penelitian yang dijalankan menghasilkan keluaran berupa sentimen masyarakat tentang citra positif seorang tokoh publik. Integrasi keilmuan ini didapat dari seorang tokoh agama yang berasal dari Pandegiling (Surabaya Pusat), yang bernama Ustad Nur Hidayat, S. Ag. Dengan melakukan wawancara pada tanggal 17 April 2019 terkait dengan hubungan antara tema yang diangkat penulis dengan sisi keislaman. Analisis sentimen masyarakat tersebut nantinya diharapkan dapat menjadi tolak ukur untuk seorang tokoh agar mampu mencerminkan citra positif yang baik sesuai harapan masyarakat.

Allah SWT berfirman dalam Qur'an surah Hujurat ayat 6 bahwa sebagai manusia hendaknya memeriksa terlebih dahulu suatu berita atau informasi yang datang kepadanya. Dengan cara menanyakan atau mencari tahu kebenaran dari berita tersebut, karena jika berita tersebut ternyata merupakan berita yang tidak benar maka dapat mencelakakan sekumpulan manusia yang lain.

"Hai orang orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti, agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu" (QS. Al-Hujurat : 6)

Dalam tafsir *fizhilalil qur'an* dijelaskan bahwa Allah memusatkan perhatian kepada orang fasik sebab dia dicurigai sebagai sumber dari suatu kebohongan. Dan agar keraguaan tidak tersebar di kalangan kaum muslimin lainnya karena berita yang disebarkaan oleh setiap individu. lalu ia menodai informasi mengubahnya atau melebihkan informasi tersebut. Pada hakekatnya, sebaliknya setiap orang dari kaum muslimin menjadi sumber informasi yang terpercaya dan hendaknya informasi tersebut benar adanya dan dapat dijadikan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pegangan. Adapun orang fasik, maka dia akan menjadi sumber dari keraguan. Dengan cara seperti itu, urusan umat akan menjadi lebih stabil, untuk bisa memilih dengan mengambil dan menolak berita yang sampai kepadanya. Kaum muslimin juga hedaknya jangan terburu-buru bertindak berdasarkan berita dari orang fasik. Karena, terburu-buru itu membuatnya bertindak zalim kepada suatu kaum sehingga akhirnya dia akan meyesali perbuatannya itu karena bertentangan dengan ajaran yang di ajarkan oleh Rasullullah dan merupakan perbuatan dimurkai Allah

Berpegang dari ayat tersebut, maka penulis mencoba membangun sebuah alat yang salah satu dari tujuannya yaitu *tabayyun* dengan cara mengumpulkan berbagai opini dari masyarakat lalu menganalisisnya sehingga mendapatkan hasil yang nantinya dapat berguna bagi masyarakat. Selain itu, dalam menyampaikan suatu pendapat atau opini sebaiknya dikalukan dengna baik dan bijaksana, akar perkataan yang keluar tidak menyakiti hati sesama.. Hal ini seperti dijelaskan dalam *qur'an* Surat an-Nisa ayat 9 berikut ini:

"Dan hendaklah takut kepada Allah orang-orang yang seandainya meninggalkan dibelakang mereka anak-anak yang lemah, yang mereka khawatir terhadap (kesejahteraan) mereka. Oleh sebab itu hendaklah mereka bertakwa kepada Allah dan hendaklah mereka mmengucapkan perkataan yang benar qaulan sadida".

Dalam surat an-Nisa ayat 9 di atas dapat diartikan sebagai sebuah ucapan yang benar dan bijaksana. Maka dari itu, untuk menyampaikan sebuah pendapat atau opini dalam kehidupan bermasyarakat sehari hari hendaknya berdasarkan fakta yang ada dan disampaikan dengan cara yang baik dan bijaksana.Sama halnya dengan menyampaikan pendapat atau opini di soaial media, Karena dengan menggunakan sosial media maka jangkauannya akan semakin luas. Ketika seseorang menyampaikan pendapat atau opini mereka di media sosial, tentu pendapat tesebut selalu berpotensi untuk diakses, dibaca, dan disebarkan oleh banyak orang. Maka dari itu, ketika menyampaikan sebuah pendapat,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

alangkah baiknya jika dilakukan berdasarkan fakta yang ada dan juga tidak menyinggung dari suatu pihak atau golongan tertentu.

Ayat tersebut telah memberikan motivasi kepada penulis untuk mencari solusi dari masalah sentimen masyarakat terkait dengan tokoh publik. Akhirnya solusi yang diperoleh untuk mempermudah analisis sentimen masyarakat terkait para tokoh publik menggunakan metode *Naïve Bayes Classifier* sesuai dengan firman Allah SWT bahwa manusia harus melakukan *tabbayun* terlebih dahulu jika mendapatkan suatu berita atau informasi dari orang lain.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 3

METODOLOGI PENELITIAN

3.1 Jenis Penelitian

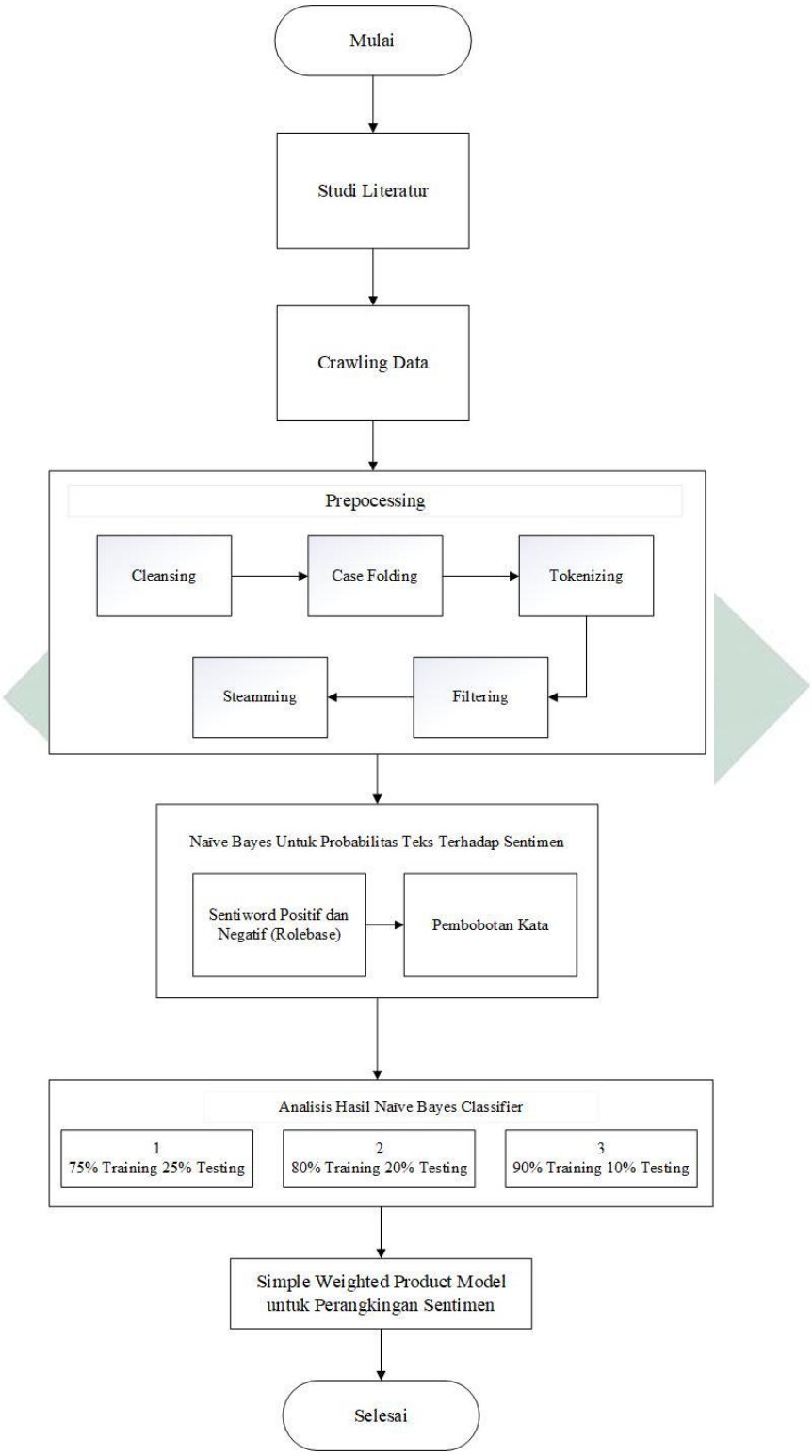
Kuantitatif deskriptif merupakan pendekatan yang digunakan dalam penelitian kali ini. Perhitungan numerik yaitu perhitungan yang digunakan dalam teori kali ini untuk mengetahui peluang dari suatu kejadian tertentu. Kejadian yang dimaksud adalah suatu opini yang termasuk dalam kategori tertentu. Selanjutnya, interpretasi dari hasil penelitian dilakukan dalam bentuk deskripsi. Oleh karena itu, jenis penelitian ini adalah penelitian kuantitatif yang bersifat deskriptif.

3.2 Jenis dan Sumber Data

Penelitiaan ini menggunakan kumpulan dari dataset tentang opini dari yang berhubungan dengan tokoh publik yang popular di tahun 2019. Data ini didapatkan dari situs www.api.twitter.com pada periode bulan Januari - Mei 2019. Untuk mendapatkan data tersebut, tahapan pertama yaitu melakukan proses *crawling data* dari alamat www.api.twitter.com dengan menggunakan aplikasi TAGSv6.1. Pada langkah ini dilakukan *crawling* dengan cara menentukan kata utama apa yang ingin dicari. Sebagai contoh, jika data yang dicari adalah opini tentang tokoh publik yang popular di tahun 2019, maka bisa dilakukan dengan menggunakan kata kunci “@Prabowo” “@Jokowi” “@Sandiuno” “@KHMarufAmin_”.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3.3 Metodologi



Gambar *3.1* Metodologi Penelitian

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Terdapat tujuh proses yang harus dilalui dalam penelitian ini jika ingin memperoleh hasil dari klasifikasi dokumen. Langkah awal yang dilakukan yaitu proses studi literatur dari berbagai macam sumber. Yang kedua yaitu dengan mengumpulkan data dengan cara *crawling* data yang dipakai untuk memperoleh dataset berupa opini masyarakat pada twitter. Langkah yang ketiga yaitu pre-proses teks yang berguna untuk merubah dataset yang belum diolah menjadi dataset yang siap untuk diklasifikasikan. Langkah yang keempat yaitu proses naïve bayes classifier pembobotan kata dalam penelitian ini menggunakan perhitungan *term frequency* (TF) dan menggunakan sentiword positif dan negatif . Perhitungan dalam pembobotan ini akan digunakan sebagai data pada penerapan metode *naïve bayes* untuk menentukan topik. Proses yang kelima yaitu analisis dari hasil klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* menggunakan tiga kali proporsi pembagian data dan dengan dua proses pengambilan data. Proses yang keenam adalah perhitungan peringkat sentimen menggunakan *Simple Weighted Product Model* sehingga didapatkan hasil peringkat klasifikasi sentimen. Agar lebih jelas, proses-proses tersebut dipaparkan seperti berikut:

3.3.1 C*rawling* Data

Untuk melakukan suatu penelitian, terlebih dahulu menyiapkan apa saja data yang dibutuhkan, maka yang perlu dilakukan yaitu proses *crawling* data dari twitter. Proses awal yang harus dikerjakan untuk melakukan *crawling* data twitter yaitu mempunyai akun twitter terlebih dahulu. Jika belum memiliki akun twitter disarankan untuk membuat akun twitter untuk memudahkan dalam proses *crawling* data. Karena sudah memiliki aku teitter maka proses selanjutnya adalah registrasi di https://dev.twitter.com/apps/new untuk memperoleh kode akses Api dari twitter. Setelah memperoleh kode akses Api dari twitter, langkah selanjutnya yaitu menggunakan *software* V Tag 6.1 untuk melakukan proses *crawling* data twitter. Tahapan *crawling* data menggunakan software V Tag 6.1 yaitu dengan cara memasukkan kata kunci yang akan dicari setelah itu menentukan waktu

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pengambilan data (biasanya hanya berskisar sampai 7 hari kebelakang) setelah itu di jalankan. Hasil dari crawling data ini terdapat delapan belas variable data yang di dapatkan. Data yang di dapatkan adalah id_str, from_user, text, created_at, time, geo_coordinates, user_lang, in_reply_to_user_id_str, in_reply_to_screen_name, from_user_id_str, in_reply_to_status_id_str, source, profile_image_url, user_followers_count, user_friends_count, user_location, status_url, entities_str. Data yang digunakan untuk langkah selanjutnya yaitu hanya data text saja. Data teks digunakan untk klasifikasii sentimen.

Tabel 3.1 Contoh Data Uji Twitter

T	Tokoh	*Tweet*	Sentimen
T1	Joko Widodo	Kesuksesan dalam pembangunan infrasruktur menghasilkan kesejahteraan, semoga sukses dan terus berhasil.	Positif
T2	Joko Widodo	Belum meratanya pembangunan menjadi bukti, bahwa gagalnya pembangunan di daerah tertinggal.	Negatif
T3	Prabowo Subianto	Ini yang diinginkan rakyat, pemimpin tegas dan dekat dengan para ulama.	Positif
T4	Prabowo Subianto	Sebuah pekerjaan jangan hanya sebatas wacana saja, tapi harus terwujud dan berhasil.	Netral
T5	Kh Ma'ruf Amin	Tutur Bahasa yang pantas dalam berbicara, logat yang tidak asal, dan bisa dipertanggungjawabkan.	Netral
T6	Sandiaga Salahudin Uno	Muda, pintar,sopan dan bijaksana.	Positif
T7	Joko Widodo	Tak pandai, bodoh, saya ragu dia pernah lulus sekolah.	Negatif
T8	Joko Widodo	Kenaikan harga pangan secara bertahap bukti	?

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		kegagalan pemerintah untuk mendengar keluhan rakyat.	

Dari Tabel 3.1 diatas bahwa terdapat tujuh teks twitter yang telah dilakukan *crawling*. Ketujuh teks tersebut terdiri dari sentimen negatif maupun positif. Data yang digunakan sebagai data latih adalah T1 sampai dengan T7 dan terdapat satu teks yang masih belum ketahui sentimennya yakni T8 . Data latih didapatkan dengan membaca teks lengkap twitter dan melakukan perbandingan untuk menentukan topik dan sentimen dengan acuan basis data *opinion word*. Cara menentukan topik dan sentimen yang digunakan sebagai data latih dengan mencocokkan kumpulan data positif dan negatif, jika pada kumpulan kata *tweet* tersebut diambil kategori sentimen yang dominan.

3.3.2 *Prepocessing*

Secara umum sistem ini memiliki beberapa proses, yang pertama adalah *prepocessing*. Tahap *prepocessing* mencangkup berbagai proses seperti *case folding, tokenizing, filtering, stemming*. Pada tahap *prepocessing* ada beberapa proses yang akan dilakukan. Pertama yakni memasukkan data inputan dimana pada tahap ini akan dimasukkan berupa teks yang berasal dari media sosial twitter. Setelah teks twitter diinputkan maka teks tersebut akan dilakukan proses *case folding* yakni teks inputan dilakukan perubahan huruf ke huruf kecil (l*owercase*) untuk semua teks. Proses *tokenzing* merupakan proses setalah *case folding*, proses tokenizing merupakan proses memecah teks yang awalnya satu kalimat dirubah menjadi perkata. Proses *filtering* yakni melakukan pembuangan kata-kata yang tidak bermakna dan menghapus kata yang muncul berkali-kali sehingga kata yang muncul berkali kali tidak mempengaruhi bobot kata tersebut. Setelah proses *filtering* maka teks inputan akan dilakukan proses selanjutnya yakni *stemming* dimana dilakukan pembuangan imbuhan pada kata sehingga menjadi kata dasar.

1. *Cleansing*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tahapan awal dalam pre-proses teks adalah *Cleansing*. Tahapan *Cleansing* ini bertujuan menghapus karakter html atau web yang ada tetapi tidak bermakna dalam proses klasifiksi dokumen. Hal ini terjadi karena dalam sebuah tweet tidak jarang masih terdapat suatu alamat web atau html yang jika tidak dihilangkan maka akan menganggu proses klasifikasi

2. *Case Folding*

Tahap ini merupakan tahap dimana teks inputan yang berasal dari sosial media twitter pertama di proses. *Case folding* adalah proses penyamaan huruf teks. Pada data *tweet* terdapat kata yang berhuruf besar dan terdapat teks yang berhuruf kecil. Perbedaan bentuk huruf pada data twitter tersebut mengganggu proses selanjutnya maka dari itu diperlukan prses *case folding* untuk merubah teks twitter menjadi huruf kecil semua seperti pada Tabel 3.2 berikut

Tabel 3.2 Teks Twitter hasil *Case Folding*

T	*Tweet*
T1	kesuksesan dalam pembangunan infrasruktur menghasilkan kesejahteraan, semoga sukses dan terus berhasil.
T2	belum meratanya pembangunan menjadi bukti, bahwa gagalnya pembangunan di daerah tertinggal.
T3	ini yang diinginkan rakyat, pemimpin tegas dan dekat dengan para ulama.
T4	sebuah pekerjaan jangan hanya sebatas wacana saja, tapi harus terwujud dan berhasil.
T5	tutur bahasa yang pantas dalam berbicara, logat yang tidak asal, dan bisa dipertanggungjawabkan.
T6	muda, pintar,sopan dan bijaksana.
T7	tak pandai, bodoh, saya ragu dia pernah lulus sekolah.
T8	kenaikan harga pangan secara bertahap bukti kegagalan pemerintah untuk mendengar keluhan rakyat.

3. Tokeninzing

Tahap *tokeninzing* meruapakn tahap yang berfungsi untuk memisahkan antara kata yang ada dalam teks twitter. Proses ini menggunakan bantuan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

karakter spasi dalam pemisah kata-nya. Pada sebuah teks twitter terdiri dari bebrapa kata yang saling terhubung dan dipisahkan dengan spasi. Untuk mempermudah dalam pemerosesan sebuah teks maka masing-masing kata dalam sebuah kalimat tersebut harus dipisahkan. Pada proses *tokeninzing* tidak hanya kata yang akan dipisahkan, angka dan tanda baca juga dipisahkan. Pemisahan tanda baca dan angka juga harus dilakukan karena mampu mempengaruhi dalam penilaian terhadap sebuah kalimat atau kata tersebut. Teks twitter pada Tabel yang sudah dilakukan *case folding* akan dilakukan *tokenizing* sehingga menjadi potongan kata seperti pada Tabel 3.3 seperti berikut:

Tabel 3.3 Term Hasil Proses *Tokeninzing*

T	*Tweet*
T1	kesuksesan | dalam | pembangunan | infrasruktur | menghasilkan | kesejahteraan | semoga | sukses | dan | terus | berhasil.
T2	belum | meratanya | pembangunan | menjadi | bukti | bahwa | gagalnya | pembangunan | di | daerah | tertinggal.
T3	ini | yang | diinginkan | rakyat | pemimpin | tegas | dan | dekat | dengan | para | ulama.
T4	sebuah | pekerjaan | jangan | hanya | sebatas | wacana | saja | tapi | harus | terwujud | dan | berhasil.
T5	tutur | bahasa | yang | pantas | dalam | berbicara | logat | yang | tidak | asal | dan | bisa | dipertanggungjawabkan.
T6	muda | pintar | sopan | dan | bijaksana.
T7	tak | pandai | bodoh | saya | ragu | dia | pernah | lulus | sekolah.
T8	Kenaikan | harga | pangan | secara | bertahap | bukti | kegagalan | pemerintah | untuk | mendengar | keluhan | rakyat.

4. *Filtering*

Kalimat yang sudah melewati proses pemisahan kata melalui tokenezing selanjutnya akan dilakukan proses *filtering*. Proses dimana akan menghilangkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

kata kata yang tidak penting atau tidak bermakna seperti kata sambung. Pada proses filtering dibutuhkan sebuah database yang menyimpan daftar sentiword yang digunakan sebagai acuan dalam penghapusan kata.

Kata hasil proses tokenezing digunakan sebagai inputan pada proses *filtering*. Setelah didapatkan kata tersebut maka tiap kata akan dilakukan pengecekan kata dengan pembanding yakni kata yang ada dalam *sentiword*. Jika kata tersebut terdapat dalam kumpulan kata *sentiword* dilakukan maka dilakukan pengecekan kembali untuk jumlah kata. Jika dalam pengecekan ditemukan jumlah kata lebih dari satu maka kata tersebut akan dilakukan pengecekan *sentiword* kembali. Hasil proses *filtering* yang didapatkan dari term pada tabel seperti pada gambar Tabel 3.4 berikut

Tabel 3.4 Term Hasil Proses *Filtering*

T	*Tweet*
T1	kesuksesan | pembangunan | infrasruktur | menghasilkan | kesejahteraan | sukses | berhasil
T2	meratanya | pembangunan | bukti | gagalnya | pembangunan | daerah | tertinggal
T3	diinginkan | rakyat | pemimpin | tegas | dekat | ulama
T4	pekerjaan | wacana | terwujud | berhasil
T5	bahasa | pantas | berbicara | logat | asal | dipertanggungjawabkan
T6	muda | pintar | sopan | bijaksana
T7	pandai | bodoh | ragu | lulus | sekolah
T8	kenaikan | harga | pangan | bertahap | bukti | kegagalan | pemerintah | mendengar | keluhan | rakyat

5. Stemming

Setelah melakuakn proses *filtering*, maka langkah selanjutnya adalah proses *stemming*. Proses *stemming* yaitu proses pencarian kata dasar terhadap sebuah kata yang telah dilakukan proses *filtering*. Pada proses ini kata akan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

diubah menjadi kata dasar sehingga kata yang memiliki imbuhan depan atau imbuhan belakang akan dihapus dan disesuaikan kata dasarnya.

Proses *stemming* yang akan mengubah bentuk kata yang awalnya bukan kata dasar diubah menjadi kata dasar menggunakan metode Nazief Adriani. Hasil proses stemming yang digunakan dengan menggunakan tabel menghasilkan seperti pada Tabel 3.5

Tabel 3.5 Term Hasil Proses *Stemming*

T	*Tweet*
T1	sukses | bangun | infrasruktur | hasil | sejahtera | sukses | hasil
T2	merata | bangun | bukti | gagal | bangun | daerah | tinggal
T3	rakyat | pemimpin | tegas | dekat | ulama
T4	kerja | wacana | wujud | hasil
T5	bahasa | pantas | bicara | logat | asal | tanggungjawab
T6	muda | pintar | sopan | bijaksana
T7	pandai | bodoh | ragu | lulus | sekolah
T8	naik | harga | pangan | tahap | bukti | gagal | pemerintah | dengar | keluhan | rakyat

3.3.3 Pembobotan Kata

Pembobotan kata dalam penelitian ini menggunakan perhitungan *term frequency* (TF). Perhitungan dalam pembobotan ini akan digunakan sebagai data pada penerapan metode *naïve bayes* untuk menentukan topik *twitter* dan penentuan hasil sentimen. Setelah dilakukan proses *Stemming*. Dalam proses kali ini, semua kata dapat diproses sehingga setiap kata yang terdapat pada dokumen tersebut memiliki bobot tersendiri. Setelah selesai, maka kumpulan dari dokumen (dokumen *corpus*) telah siap untuk segera dilakukan proses *training* untuk proses klasifikasi.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3.3.4 Naïve Bayes untuk Probabilitas Sentimen

Probabilitas sentimen merupakan proses penentuan teks twitter yang dihitung untuk didapatkan nilai tertinggi dari beberapa kategori dalam sentimen. Kedelapan teks yang digunakan untuk perhitungan probabilitas sentimen.

Positif : Sukses (2), bangun (1), Infrastuktur (1), hasil (2), sejahtera (1), Rakyat (1). pemimpin (1), tegas (1), dekat (1), Ulama (1), Muda (1), pintar (1), sopan (1), bijaksana (1)

Negatif : Merata (1), bangun (2), bukti (1), gagal (1), daerah (1), tinggal (1),Pandai (1), bodoh (1), ragu (1), lulus (1), sekolah (1)

Netral : kerja (1), wacana (1), terwujud (1), berhasil (1), bahasa (1), pantas (1), bicara (1), logat (1), asal (1), tanggungjawab (1)

? : naik (1), harga (1), pangan (1), tahap (1), bukti (1), gagal (1), pemerintah (1), dengar (1), keluhan (1), rakyat (1)

Seperti halnya untuk menghitung tweet terhadap topik, perhitungan probabilitas *tweet* terhadap sentimen juga menggunakan rumus 3 dan 4 dimana untuk menentukan probabilitas sentimen memerlukan nlai *p(ci)* dan *p(wkj.ci)* untuk probabilitas awal sentimen didapatkan dari jumlah tweet yang tergabung dalam topik *ci* dibagi dengan semua topik. Dari tabel (3.1) yang merupakan tweet awal, didapatkan bahwa ada tujuh (7) jumlah keseluruhan tweet data uji yang terdiri dari tiga sentimen positif, dua untuk sentimen negatif dan dua untuk sentimen netral.

p(positif) = 3/7

p(negatif) = 2/7

p(netral) = 2/7

Nilai *p(wkj.ci)* pada pembobotan sentimen dilakukan dengan rumus 4 dimana fitur kemunculan term ditambahkan satu lalu dibagi dengan hasil penjumlahan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

antara semua jumlah term pada topik ci ditambah dengan keseluruhan term. Diketahui bahwa pada sentimen positif terdapat 20 kali fitur kemunculan kata pada sentimen positif, 15 pada sentimen negatif dan jumlah keseluruhan term adalah 35.

$$p(sukses,positif) = \frac{2+1}{14+35} = \frac{3}{49}$$

$$p(sukses,negatif) = \frac{0+1}{11+35} = \frac{1}{46}$$

$$p(sukses,netral) = \frac{0+1}{10+35} = \frac{1}{45}$$

Tabel 3.6 Bobot Probabilitas *Naive Bayes* Untuk Sentimen

Topik	*p(ci)*	*p(wkj.ci)*									
		Sukses	Bangun	Inftrastruktur	Hasil	Sejahtera	Merata	Bukti	Gagal	Daerah	Tinggal
Positif	3/7	3/49	2/49	2/49	3/49	2/49	1/49	1/49	1/49	1/49	1/49
Negatif	2/7	1/46	3/46	1/46	1/46	1/46	2/46	2/46	2/46	2/46	2/46
Netral	2/7	1/45	1/45	1/45	2/45	1/45	1/45	1/45	1/45	1/45	1/45

Topik	*p(ci)*	*p(wkj.ci)*										
		Rakyat	Pemimpin	Tegas	Dekat	Ulama	Muda	Pintar	Sopan	Bijaksana	Kerja	Wacana
Positif	3/7	2/49	2/49	2/49	2/49	2/49	2/49	2/49	2/49	2/49	1/49	1/49
Negatif	2/7	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46	1/46
Netral	2/7	1/45	1/45	1/45	1/45	1/45	1/45	1/45	1/45	1/45	2/45	2/45

Topik	*p(ci)*	*p(wkj.ci)*											
		Wujud	Bahasa	Pantas	Bicara	Logat	Asal	Tangungjawab	Pandai	Bodoh	Ragu	Llulus	Sekolah
Positif	3/7	1/49	1/49	1/49	1/49	1/49	1/49	1/49	1/49	1/49	1/49	1/49	1/49
Negatif	2/7	1/46	1/46	1/46	1/46	1/46	1/46	1/46	2/46	2/46	2/46	2/46	2/46
Netral	2/7	2/45	2/45	2/45	2/45	2/45	2/45	2/45	1/45	1/45	1/45	1/45	1/45

p(“positif”|”T8”) = p(“positif”) x p(“naik”|”positif”) + p(“harga”|”positif”) + p(“pangan”|”positif”) + p(“tahap”|”positif”) + p(“bukti”|”positif”) + p(“gagal”|”positifr”) + p(“pemerintah”|”positif”) + p(“dengar”|”positif”) + p(“keluhan”|”positif”) + p(“rakyat”|”positif”)

= 3/7 x (1/49 + 1/49 + 1/49 + 1/49 + 1/49 + 1/49 + 1/49 + 1/49 + 1/49 + 2/49)

= 0,09621

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

p("negatif"|"T8") = p("negatif") x p("naik"|" negatif") + p("harga"|" negatif") + p("pangan"|" negatif") + p("tahap"|" negatif") + p("bukti"|" negatif") + p("gagal"|" negatif") + p("pemerintah"|" negatif") + p("dengar"|" negatif") + p("keluhan"|" negatif") + p("rakyat"|" negatif")

= 2/7 x (1/46 + 1/46 + 1/46 + 1/46 + 2/46 + 2/46 + 1/46 + 1/46+ 1/46+ 1/46)
= 0,074534

p("netral"|"T8") = p("netral") x p("naik"|" netral") + p("harga"|" netral") + p("pangan"|" netral") + p("tahap"|" netral") + p("bukti"|" netral ") + p("gagal"|" netral") + p("pemerintah"|" netral") + p("dengar"|" netral") + p("keluhan"|" netral ") + p("rakyat"|" netral")

= 2/7 x (1/45 + 1/45 + 1/45+ 1/45+ 1/45+ 1/45+ 1/45+ 1/45+ 1/45+ 1/45)

= 0,063492

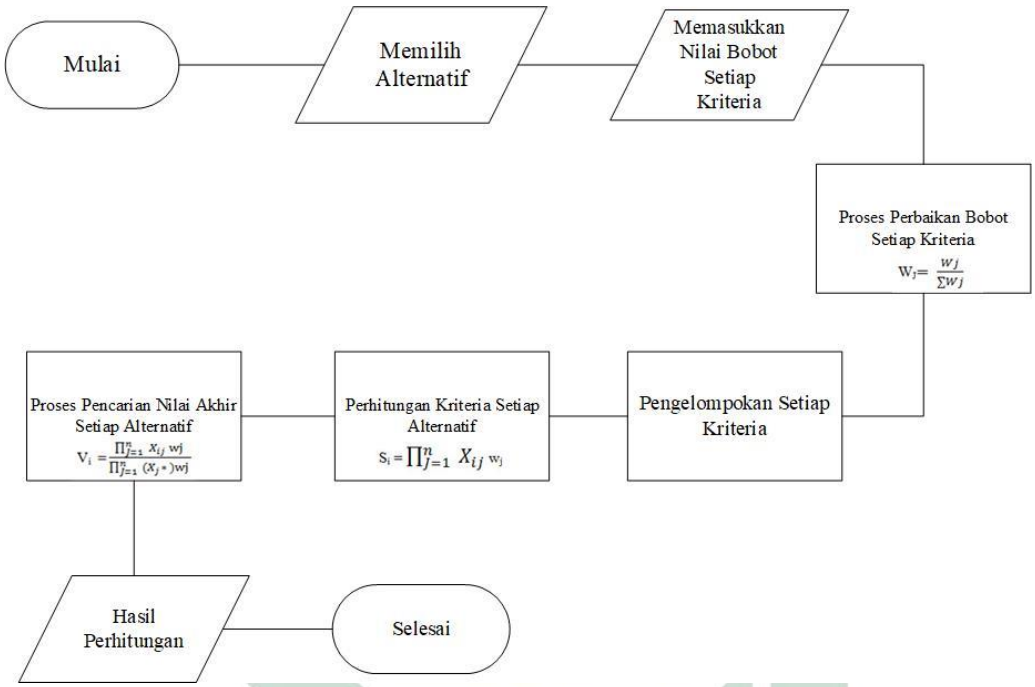
Dengan melakukan perhitungan probabilitas teks terhadap sentimen dengan menggunakan metode *naïve bayes* maka dapat diketahui bahwa teks twitter T8 termasuk dalam kategori sentimen positif. Nilai bobot probabilistic teks twitter T8 terhadap topik positif lebih besar dari pada nilai bobot teks T8 terhadap topik lainnya. Nantinya setelah data secara keseluruhan telah diolah maka siapa tokoh publik yang memiliki presentase sentimen positif terbesar maka tokoh tersebutlah yang mendapatkan citra positif teratas.

3.3.5. Proses Metode Simple Weighted Product Model

Pada penelitian ini pengunaan metode *Simple Weighted Product Model* digunakan untuk menghitung nilai sentimen yang ada pada topik setiap tokoh publik sehingga didapatkan urutan-urutan nama tokoh publik yang mendapatkan nilai sentimen terbaik sampai dengan sentimen terburuk. Langkah awal yaitu dengan memilih alternatif, kemudian memasukkan nilai bobot kepentingan dari setiap kriteria untuk mendapatkan hasil rangking. Pertama proses perbaikan bobot ini dilakukan untuk merubah nilai total bobot menjadi 1. Kemudian proses perhitungan semua kriteria yang telah dipilih akan dikelompokkan berdasarkan tipe keuntungan (max) dan (min). Kemudian menghitung nilai kriteria setiap alternatif berdasarkan rumus yang sudah ditentukan. Selanjutnya proses proses perangkingan dilakukan setelah melakukan perhitungan preferensi dengan membagi hasil vektor dari setiap kriteria dengan jumlah vektor kriteria.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Nilai akhir dari perhitungan akan digunakan untuk perangkingan. Semakin besar nilai akhir suatu alternatif (tokoh publik) maka akan mendapatkan citra positif terbaik dari masing masing tokoh publik. Proses perhitungan ini dapat dilihat pada Gambar 3.2



Gambar 3.2. Proses Perhitungan Metode *Simple Weighted Product Model*

3.3.6 Analisis Hasil

Setelah didapatkan hasil dari proses klasifikasi sentimen masyarakat terkaiat dengan citra positif tokoh publik, dan setelah pembobotan setiap kata sudah selesai dan menghasilkan pembobotan kata terhadap sentimen positif dan negative dari setiap tokoh. Data tersebutlah yang digunakan untuk pembuatan *wordcloud*. Tujuan dari pembuatan *wordcloud* ini yaitu untuk mengetahui kata kata mana saja yang sering muncul dan berpengaruh besar dalam pembuatan model klasifikasi sistem. Setelah itu melakukan probabilitas sentimen menggunakan Naïve Bayes Classifier dengan tiga kali proporsi dan dengan dua acara pengambilan data yang berbeda. Setalah itu dilakukan perhitungan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menggunakan *Simple Weighted Product Model* untuk mengetahui urutan atau peringkat dari masing masing tokoh publik yang bersangkutan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 4

HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan bagaimana proses dan hasil yang didapatkan dari penelitian kali ini mulai dari proses awal hingga proses akhir.

4.1. *Crawling* Data Twitter

Proses awal yang dikerjakan pada penelitian ini adalah *crawling* data dari twitter. *Crawling* data twitter ini dikerjakan secara bertahap. Hal ini karena data yang bisa didapatkan dari proses *crawling* terrbatas hanya tujuh hari terakhir dari waktu pencarian. Pada tahapan ini, pencarian data twitter dibagi menjadi empat bagian. Bagian yang pertama yaitu mencari data yang memiliki hubungan dengan Bapak Joko Widodo. Yang kedua yaitu mencari data yang memiliki hubungan dengan Bapak KH. Ma'ruf Amin. Yang ketiga yaitu mencari data yang memiliki hubungan dengan Bapak Prabwo Subianto. Dan yang keempat yaitu mencari data yang memiliki hubungan dengan Bapak Sandiaga Salahudin Uno. Keempat tokoh tersebut merupakan tokoh yang paling populer di Indonesia pada tahun 2019. Dimana keempat tokoh tersebut pada tahun 2019 mencalonkan dirinya untuk menjadi presiden dan wakil presiden Republik Indonesia pada periode tahun 2019-2024.

Crawling data twitter ini dilakukan pada bulan Januari-Mei 2019, Untuk mencari data yang berhubungan dengan para tokoh publik tersebut menggunakan berbagai kata kunci. Kata kunci ini merupakan nama akun twitter dari pada para tokoh publik yang bersangkutan. Yang Pertama untuk mencari kata kunci dari Bapak Joko Widodo, digunakan kata utama yaitu "@Jokowi". Sehingga nantinya akan termuat seluruh data tweet yang memiliki hubungan dengan Bapak Joko Widodo dalam tujuh hari terakhir. Dari proses crawling untuk mencari data yang memiliki hubungan dengan Bapak Joko Widodo, didpatkan data sebanyak 14.526 data. Data tersebut belum melalui tahapan penghapusan tweet yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memiliki kalimat ganda (duplikat). Setelah melalui tahapan penghapusan data maka data total yang diperoleh sebanyak 10.981. Berikut merupakan Gambar 4.1 yang merupakan contoh hasil dari *crawling* data twitter yang memiliki hubungan dengan Bapak Joko Widodo yang disimpan dalam file Microsoft excel berformat. csv (Comma Delimited).

1	id_str	from_user	text
2	1167225415712	amircunding	RT @MN_aliakbar: Yang saya hormati Pak President RI @jokowi dan Pak Gubernur Jabar @ridwankamil , Saya akan menjadi perwakilan Indonesia dan Eropa untuk mengikuti kompetisi tingkat dunia di ajang student paper contest 2019 di bidang oil and gas di SPE ATCE 2019. Tetapi sy belum memiliki sponsor .
3	1167225410306	MuhammadAndre1_	RT @jokowi: Sebanyak 3.800 lembar sertifikat hak atas tanah untuk masyarakat di Kabupaten Purworejo, Provinsi Jawa Tengah saya serahkan hari ini di GOR W.R. Supratman. Insya Allah, tahun 2023 mendatang seluruh bidang tanah di Jawa Tengah telah memiliki sertifikat. https://t.co/JF0BUCQMPi
4	116722540517	DimedjaR	RT @Dahnilanzar: Saya berkeyakinan Pak @jokowi mampu meredam gejolak di Papua dan Papua Barat. Apalagi beliau di pilih oleh lebih 90% rakyat Papua, dan sdh bnyk membangun infrastruktur disana, pastilah beliau sangat dihormati rakyat Papua.Pak Jokowi hadir disana pasti mampu meredam gejolak.
5	1167225397816	LorensiusPurba	RT @mas__piyuuu: Disaat Papua bergolak, @jokowi terpingkal-pingkal dihibur pelawak... https://t.co/vnCZVNFt6v https://t.co/yD4IKC1trQ
6	1167225395526	agunkkurniaone1	@andi_chairil @jokowi Ga usah ngomong ibukota bro papua sdg membara skrg
7	1167225391826	ronalfausen2	RT @rahung: 20 thn lalu Timor-Timur melakukan referendum stlh okupasi yg menewaskan 1/3 jmlh penduduk Timor-Leste. Menurut Pengadilan khusus PBB, yg paling bertanggungjawab atas bumi hangus Timor-Leste thn 1999 adlh Jenderal Wiranto. Namun di masa Orde Baik @jokowi diberi mandat urus Papua?
8	1167225389356	yunari_atiss	RT @ZAEffendy: #Jayapura membara pak @jokowi, pintu gerbang NKRI sudah terbuka lebar, berilah komentar yang menenangkan, bahkan perlukan terbang ke wilayah #Papua malam ini juga. Have a sense of crisis, at least? https://t.co/BAq7PQ9cgS

Gambar 4.1. Contoh Data untuk Joko Widodo

Pada gambar 4.1 dapat terlihat bahwa hasil dari proses crawling data terdapat delapan belas variable data yang di dapatkan. Data yang di dapatkan adalah id_str, from_user, text, created_at, time, geo_coordinates, user_lang, in_reply_to_user_id_str, in_reply_to_screen_name, from_user_id_str, in_reply_to_status_id_str, source, profile_image_url, user_followers_count, user_friends_count, user_location, status_url, entities_str. Data yang digunakan untuk langkah selanjutnya yaitu hanya data text saja. Data teks digunakan untk klasifikasii sentimen.

Yang Kedua, untuk mencari data yang berhubungan dengan Bapak KH. Maruf Amin digunakan kata utama yaitu "@KHMarufAmin_", sehingga nanti akan tersedia semua data yang berhubungan dengan Bapak KH. Makruf Amin

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

selama tujuh hari terakhir. Dari proses crawliing untuk mencari data yang memiliki hubungan dengan Bapak KH. Makruf Amin, didpatkan data sebanyak 13.207 data. Data tersebut belum melalui tahapan penghapusan tweet yang memiliki kalimat ganda (duplikat). Setelah melalui tahapan penghapusan data maka data total yang diperoleh sebanyak 11.351. Berikut adalah Gambar 4.2 yang merupakan contoh hasil dari proses *crawling* untuk KH.Makruf Amin.

1	id_str	from_user	text
2	1167087	r_cho888	Karena media cuma heboh beritakan soal wacana Pindah Ibu Kota & Iuran BPJS Naik, ternyata banyak rakyat tidak tahu mengenai #PapuaBergejolak. Mau sampai kapan? Katanya 90% memilih Paslon 01 @jokowi @KHMarufAmin_, tapi kenapa rakyat Papua lebih memilih referendum? #WeLovePapua
3	116706	B_Hendratmo1	@snjayajonggrang Kenapa @KPI_Pusat TIDAK beri sanksi terhadap @ILCtv1 jika benar ucapan Ridwan Saidi (budayawan Betawi) hoax ? CC YTh Bapak-Bapak @jokowi @KHMarufAmin_ @wiranto1947 @GeneralMoeldoko @DivHumas_Polri @BadjaNuswantara @KanjengRaden_ @indonesiamegah @liriksoe @permadiaktivis
4	116658	AJudakusumah	@AmbarwatiRexy Bgmn kalau gubernurnya kita mintakan klarifikasi ? 🙏 @jokowi @Pak_JK @KHMarufAmin_ @pramonoanung @GeneralMoeldoko @KSPgoid @Wahyu_MetroTV @AimanWitjaksono @Fitri_MetroTV @Bayusutiyono @budiadiputro_ @NajwaShihab
5	116655	AJudakusumah	@edhi_shi @SamSoegi @CoretPdipilpres @PartBurungGalak Betul juga, ada beberapa sinetron bagus yg sayang entah kenapa hanya tinggal kenangan. Kurang komersil ? Pdhal sarana pencerahan rakyat juga. @KSPgoid @puan_maharani @imam_nahrawi @YohanaYembise @muhadjir_ef @KHMarufAmin_

Gambar 4.2. Contoh Data Untuk KH.Makruf Amin

Pada gambar 4.2 dapat dilihat bahwa dari hasil proses *crawling* data dari KH.Ma'ruf Amin. Data yang digunakan untuk langkah selanjutnya yaitu hanya data text saja. Data teks dipakai untk klasifikasi sentimen.

Yang Ketiga, untuk mencari data yang berhubungan dengan Bapak Prabowo Subianto digunakan kata utama "@Prabowo", sehingga nanti akan tersedia semua data yang memiliki hubungan dengan Bapak Prabowo Subianto selama tujuh hari terakhir. Dari proses crawling untuk mencari data yang berhubungan dengan Bapak Prabowo Subianto, didpatkan data sebanyak 17.003 data. Data tersebut belum melalui tahapan penghapusan *tweet* yang memiliki kalimat ganda (duplikat). Setelah melalui tahapan penghapusan data maka data

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

total yang diperoleh sebanyak 15.600. Berikut adalah Gambar 4.3 yang merupakan contoh hasil dari proses crawling data untuk Prabowo Subianto.

1	id_str	from_user	text
2	116722	pasadenasl	RT @dr_koko28: Bukan tugas utama pak @prabowo atau ustad Abdul Somad atau Aa Gym untuk mendamaikan dan menormalkan kembali kehidupan masyarakat kita di Papua. Tunggu,... atau apakah sebenarnya iya?🤔 #WeLovePapua
3	116722	yosgabe4	@CeciliaSuwanda1 @adeira78 @Dahnilanzar @jokowi @prabowo Masyarakat Papua sebagian besar gak kenal Wowo, sebagian lainnya ingat sejarah Mapenduma saat Wowo bantai warga sipil yg dituduh OPM. Warga Papua juga Tidak lupa, bapaknya Wowo dan Suharto yg di 1967 mulai MENJUAL Freeport. Itulah sebab 2014 & 2019 Wowo KALAH TELAK di Papua
4	116722	evalez_fr	@Jsubagja7 @RajaGuguk14 @Rasyeed3 @Miqdad_2 @ekosubahri1 @lubis_mariska @prabowo @rockygerung @NMoekijat @Dahnilanzar @probowo @Ronnie_Rusli @zarazettirazr @haikal_hassan Iya sya juga lg nyantai, ga ada kerjaan malah disuruh ikut diklat/seminar terus wkwk. Ya ngga apa² lah di ongkosin ini 😐😐
5	116722	SorosMas	RT @dr_koko28: Bukan tugas utama pak @prabowo atau ustad Abdul Somad atau Aa Gym untuk mendamaikan dan menormalkan kembali kehidupan masyarakat kita di Papua. Tunggu,... atau apakah sebenarnya iya?🤔 #WeLovePapua
6	116722	sang_pnakluk	@alfaqir_ponco @prabowo @petruci_idn Na'am Sheikh
7	116722	sang_pnakluk	RT @alfaqir_ponco: @sang_pnakluk @prabowo Afwan Bang...lanjutin amanah, ajun sohib kite @petruci_idn pemerhati Palestina akan tumbang🙏

Gambar 4.3 Contoh Data Untuk Prabowo Subianto

Pada gambar 4.3 dapat dilihat bahwa dari hasil crawling data dari Prabowo Subianto. Data yang dipakaiuntuk proses selanjutnya yaitu hanya data text saja. Data teks digunakan untk klasifikasi sentimen.

Dan yang terakhir, untuk mencari data yang memiliki hubungan dengan Bapak Sandiaga Salahudin Uno digunakan kata utama yaitu "@Sandiuno", sehingga nanti akan muncul semua data yang berhubungan dengan Bapak Sandiaga Salahudin Uno selama tujuh hari terakhir. Dari proses crawling untuk mencari data yang memiliki hubungan dengan Bapak Sandiaga Salahudin Uno, didapatkan data sebanyak 12.533 data. Data tersebut belum melalui tahapan penghapusan tweet yang memiliki kalimat ganda (duplikat). Setelah melalui tahapan penghapusan data maka data total yang diperoleh sebanyak 11.687. Berikut adalah Gambar 4.4 yang merupakan contoh hasil dari proses crawling data untuk Sandiaga Salahudin Uno.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1	id_str	from_user	text
2	116960	Abd7IK	RT @sandiuno: Siang tadi saya mengunjungi Panti Sosial Bina Laras Harapan Sentosa 1 Cengkareng,Jakarta Barat. Di tempat ini warga binaan sosial dengan gangguan kejiwaan dibina & dilatih untuk menghasilkan sebuah karya yang diharap dapat membawa manfaat untuk dirinya sendiri & untuk orang lain. https://t.co/KGTyrexAWC
3	116960	GlowNop	RT @nurasiauno: Berbuat baik bisa kapan saja, di mana saja, tidak perlu dengan kekuasaan. Kalau saya inginnya Mpok @nurasiauno jadi ibu rumah tangga dan tetap menjadi Mamah Online saja, pesan tersirat dari bang @sandiuno , baiklah papa online 😂😂😍😍 https://t.co/uxXq8IT3IO
4	116960	pebriana_wahid	RT @sandiuno: Siang tadi saya mengunjungi Panti Sosial Bina Laras Harapan Sentosa 1 Cengkareng,Jakarta Barat. Di tempat ini warga binaan sosial dengan gangguan kejiwaan dibina & dilatih untuk menghasilkan sebuah karya yang diharap dapat membawa manfaat untuk dirinya sendiri & untuk orang lain. https://t.co/KGTyrexAWC
5	116960	KacungBukan	RT @sandiuno: Siang tadi saya mengunjungi Panti Sosial Bina Laras Harapan Sentosa 1 Cengkareng,Jakarta Barat. Di tempat ini warga binaan sosial dengan gangguan kejiwaan dibina & dilatih untuk menghasilkan sebuah karya yang diharap dapat membawa manfaat untuk dirinya sendiri & untuk orang lain. https://t.co/KGTyrexAWC
6	116960	harry52541661	RT @sandiuno: Siang tadi saya mengunjungi Panti Sosial Bina Laras Harapan Sentosa 1 Cengkareng,Jakarta Barat. Di tempat ini warga binaan sosial dengan gangguan kejiwaan dibina & dilatih untuk menghasilkan sebuah karya yang diharap dapat membawa manfaat untuk dirinya sendiri & untuk orang lain. https://t.co/KGTyrexAWC
7	116960	KarimTukih	RT @sandiuno: Siang tadi saya mengunjungi Panti Sosial Bina Laras Harapan Sentosa 1 Cengkareng,Jakarta Barat. Di tempat ini warga binaan sosial dengan gangguan kejiwaan dibina & dilatih untuk menghasilkan sebuah karya yang diharap dapat membawa manfaat untuk dirinya sendiri & untuk orang lain. https://t.co/KGTyrexAWC

Gambar 4.4. Contoh Data Untuk Sandiaga Salahudin Uno

Pada gambar 4.4 dapat dilihat bahwa dari hasil crawling data dari Sandiaga Salahudin Uno. Data yang dipakai untuk langkah selanjutnya yaitu hanya data text saja. Data teks dipakai untk klasifikasi sentimen.

4.2. *Prepocessing*

Data *tweet* tokoh publik yang teridi dari keempat tokoh sudah diambil melalui twitter API dengan menggunakan aplikasi TAGSv6.1, selanjutnya akan dilakukan proses *preprocessing.* Tahap *prepocessing* mencangkup berbagai proses seperti *cleansing, case folding, tokenizing, filtering, stemming*. Pada tahap *prepocessing* ada beberapa proses yang akan dilakukan. Pertama yakni memasukkan data inputan dimana pada tahap ini akan dimasukkan berupa teks yang berasal dari media sosial twitter. Setelah teks twitter diinputkan maka teks tersebut akan dilakukan proses cleansing yaitu untuk menghapus semua karakter halaman html atau web yang tidak bermakna dalam mengklasifikasikan dokumen. Selanjutnya yakni proses *case folding* yakni teks inputan dilakukan perubahan huruf ke huruf kecil (*lowercase*) untuk semua teks. Proses *tokeninzing* merupakan proses setalah *case folding*, proses *tokenizing* merupakan proses

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memecah teks yang awalnya satu kalimat dirubah menjadi perkata. Proses filtering yakni melakukan pembuangan kata-kata yang tidak bermakna dan menghapus kata yang muncul berkali-kali sehingga kata yang muncul berkali kali tidak mempengaruhi bobot kata tersebut. Setelah proses *filtering* maka teks inputan akan dilakukan proses selanjutnya yakni *stemming* dimana dilakukan pembuangan imbuhan pada kata sehingga menjadi kata dasar.

```
#prepocessing
fPreprocess = function(data){
  data = tm_map(data, fSpace, "/")
  data = tm_map(data, fSpace, "@")
  data = tm_map(data, fSpace, "\\|")
  data = tm_map(data, fSpace, "[[:punct:]]")
  data = tm_map(data, fSpace, "[[:digit:]]")
  data = tm_map(data, content_transformer(tolower))
  data = tm_map(data, fUrl)
  data = tm_map(data, stripWhitespace)

  data = paste0(data)
  data = tokenize_words(data)
  data = Corpus (VectorSource(data))

  data = tm_map(data, fSpace, '"')
  data = tm_map(data, fSpace, ',')
  data = tm_map(data, removeWords, c('list','language','en'))

  return(data)
}
```

Gambar 4.5. *Source Code* Proses Prepocessing Data

4.3 Manual dan *Rulebase Clasification*

Penelitian ini menggunakan *rulebase* yaitu sentiword positif dan negatif Bahasa Indonesia yang di dapatkan dari Github. Alasan digunakannya rulebase ini adalah untuk efensiensi waktu dalam melakukan klasifikasi data twitter, mengingat jumlah data twitter yang dijadikan sampel lebih dari 10.000 data twit. Peneliti juga pernah melakukan klasifikasi data twitter secara manual. Tetapi dalam jangka waktu 7 hari peneliti hanya berhasil mengklasifikasikan teks twitter sebanyak 800 teks twitter saja. Oleh karena itu peneliti menggunakan *rulebase* agar lebih menghemat waktu pengerjaan proses klasifikasi.

4.4. Pembobotan Kata

Langkah selanjutnya yaitu pembobotan kata, pembobotan kata dalam penelitian ini menggunakan perhitungan *term frequency* (TF). Perhitungan dalam pembobotan ini akan digunakan sebagai data pada penerapan metode naïve bayes

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

untuk menentukan topik twitter dan penentuan hasil sentimen. Adapun dasar dari pembobotan kata ini menggunakan suatu *rulebase* (sentiword) positif dan negatif yang di dapat dari Github. Perhitungan term atau kata menggunakan kata hasil *preprocessing* dan menghasilkan bobot masing-masing kata seperti pada Tabel dibawah ini:

4.4.1 Contoh Pembobotan Kata untuk Joko Widodo

Contoh pembobotan kata untuk tokoh publik yang pertama yaitu Bapak Joko Widodo, bisa dilihat pada gambar 4.6 bahwa hasil dari pembobotan kata untuk tokoh publik Bapak Joko Widodo adalah sebagai berikut:

	positif	negatif	hasil
1	1	0	positif
2	0	1	negatif
3	0	1	negatif
4	1	2	negatif
5	2	3	negatif
6	8	0	positif
7	2	1	positif
8	2	0	positif
9	4	3	positif
10	1	4	negatif
11	3	1	positif
12	0	1	negatif
13	1	0	positif
14	0	1	negatif
15	0	1	negatif
16	1	2	negatif
17	4	2	positif
18	1	0	positif
19	2	0	positif
20	1	2	negatif
21	2	0	positif
22	2	0	positif

Showing 1 to 23 of 5,755 entries

Gambar 4.6. Pembobotan Kata untuk Joko Widodo

Tidak hanya itu pembobotan kata adalah metode yang sederhana karena dari setiap kata ditentukan dari banyaknya kemunculan kata yang ada pada *tweet* yang bersangkutan. Hal ini juga menjadi bahan untuk pembuatan *wordcloud* sentimen positif dan negatif dari tokoh publik yang bersangkutan. *Wordcloud* ini dibuat untuk mengetahui kata sentimen positif dan negatif mana yang paling sering dipakai masayarakat untuk menilai tokoh publik yang mereka perbincangkan. Seperti pada gambar 4.7 di bawah ini merupakan bobot probabilitas kata sentimen positif dari Joko Widodo.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	word	freq
banyak	banyak	320
kerja	kerja	308
semoga	semoga	289
baik	baik	285
menang	menang	223
selamat	selamat	216
maju	maju	164
kasih	kasih	162
bersama	bersama	148
jelas	jelas	133
sehat	sehat	132
kemenangan	kemenangan	118
suka	suka	114
siap	siap	109
mantap	mantap	101
semangat	semangat	101
benar	benar	97
percaya	percaya	97
membangun	membangun	90
penting	penting	90
setuju	setuju	90
yakin	yakin	85

Showing 1 to 23 of 1,884 entries

Gambar 4.7. Bobot Probabilitas Kata Sentimen Positif untuk Joko Widodo

Dapat dilihat dari gambar diatas bahwa bobot probabilitas kata sentimen positif yang paling sering digunakan masyarakat dalam memerikan tweet atau opini positif sesuai dengan kamus sentimen positif yang didapat dari Github untuk Joko Widodo. Terdapat 1.884 baris kata yang sering digunakan masyarakat yang pertama yaitu kata “banyak” yang digunkan sebanyak 320 kali. Setelah itu kata “kerja” yang digunakan sebanyak 308 kali, dan seterusnya. Berikut ini merupakan *wordcloud* dari sentimen positif yang didapat oleh Joko Widodo:

Gambar 4.8. *Woldcloud* sentimen positif Joko Widodo

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	word	freq
kalah	kalah	132
kecurangan	kecurangan	94
curang	curang	85
fitnah	fitnah	84
lupa	lupa	84
masalah	masalah	81
padahal	padahal	66
mahal	mahal	60
dungu	dungu	56
gila	gila	56
kasus	kasus	54

korban	korban	53
asal	asal	45
lama	lama	45
goblok	goblok	43
hutang	hutang	43
korupsi	korupsi	41
harusnya	harusnya	40
gagal	gagal	39
bodoh	bodoh	38
bohong	bohong	37
nyinyir	nyinyir	37

Showing 1 to 23 of 2,056 entries

Gambar 4.9. Bobot Probabilitas Kata Sentimen Negatif untuk Joko Widodo

Dapat dilihat dari gambar diatas bahwa bobot probabilitas kata sentimen negatif yang paling sering digunakan masyarakat dalam memerikan tweet atau opini negatif sesuai dengan kamus sentimen negatif yang didapat dari Github untuk Joko Widodo. Terdapat 2056 baris kata yang sering digunakan masyarakat yang pertama yaitu kata “kalah” yang digunkan sebanyak 132 kali. Setelah itu kata “kecurangan” yang digunakan sebanyak 94 kali, dan seterusnya. Berikut ini pada gambar 4.11 merupakan *wordcloud* dari sentimen negatif yang didapat oleh Joko Widodo:

Gambar 4.10. *Woldcloud* Sentimen Negatif Joko Widodo

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.5. *Naïve Bayes* Untuk Probabilitas Teks Terhadap Sentimen

Setelah semua kata yang terdapat pada data set telah diberi bobot untuk masing masing tokoh publik, maka langkah selanjutnya yaitu dengan membuat sistem klasifikasai menggunakan metode *Naïve Bayes Clasiffier* seperti yang telah dipaparkan pada bab dua mengenai bagaimana cara kerja dari *Naïve Bayes Classifier* berdasarkan pada peluang dari kemunculan kata yang telah diberikan bobot pada sebelumnya.

Setelah dilakukannya pelabelan data sesuai dengan *rulebase* yang ada, selanjutnya yaitu melakukan penghapusan data netral yang tidak ada kaitannya dengan penelitian kali ini. Proses menghilangkan data netral ini dilakukan karena keberadaan data netral yang cukup banyak, sehingga akan mengganggu dari proses klasifikasi. Selain itu data yang dipakai sebagai data set hanya sebatas data yang berlabel sentimen positif dan negatif saja. Salah satu contoh dari data ntral yaitu data yang tidak termasuk ke dalam sentiword negative dan sentiword positif, adapun data netral ini yaitu data terdiri dari kata yang berbahasa Inggris, kata yang tidak sesuai dengan ejaan yang tepat (typo), dan kata kata singkatan.

Pada sistem klasifikasi, tahapan awal yang dilakukan adalah dengan membagi data menjadi dua bagian yaitu data training dan data testing. Pembagian data ini akan dilakukan secara acak dan urut, hal ini dilakukan karena untuk mengetahui tingkat akurasi yang dihasilkan dari kedua cara tersebut mana yang menghasilkan akurasi yang lebih tinggi dari keempat calon publik yang bersangkutan. Pada penelitian kali ini peneliti akan membagi data menjadi 3 kali proporsi.

1. 75% data training dan 25% data testing
2. 80% data training dan 20% data testing
3. 90% data training dan 10% data testing

Setelah proses sistem klasifikasi berakhir dari data training yang tersedia. Proses selanjutnya yang harus dilakukan adalah pengujian terhadap sistem klasifikasi

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang dibuat. Evaluasi merupakan tahapan yang paling menentukan, karena bisa dilihat dari proses evaluasi tersebut apakah sistem klasifikasi yang sudah dibuat mempunyai peforma yang baik atau sebaliknya. *Confusion matrix* adalah salah satu cara untuk melakukan pengujian pada sistem klasifikasi. *Confusion Matrix* juga bisa dipakai untuk menghitunng tingkat akurasi, presisisi, dan recall. Perhitungan evaluasi sistem klasifikasii pada penelitian ini terbagi menjadi empat bagian. Hal ini dikarenakann sistem yang di buat ada empat yang masing masing untuk para tokoh publik yang ada.

4.5.1 Skenario Proporsi Pertama 75 % Training 25 % Testing

1. Joko Widodo

Pada sistem yang pertama yaitu untuk tokoh publik Joko Widodo data yang digunakan berjumlah 10.981 data. Seltelah melaui proses penghapusan data netral maka data yang diperoleh oleh tokoh publik Joko Widodo adalah 5.755 data. Pada pembagian proporsi data yang pertama yaitu 75% data training dan 25% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4316 data dan data testing sebanyak 1439 data.

Proporsi Pertama Urut

	testings	Freq
1	False Negatif	24
2	True Negatif	402
3	True Positif	1013

Gambar 4.11 *Confusion Matrix* Proporsi Pertama Urut

Gambar 4.11 Menunjukkan *Confusion Matrix* untuk pengambilan data secara urut pada proporsi pertama

Tabel 4.1. *Confusion Matrix* Proporsi Pertama Urut

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	1013	24	1037
Negatif	0	402	402
Jumlah Kolom	1013	426	1439

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi pertama urut pada tokoh publik Joko Widodo

Akurasi = $\frac{1013+402}{1013+402+0+24} . 100\%$

= 98,33 %

Presisi Positif = $\frac{1013}{1013+0} . 100\%$

= 100 %

Presisi Negatif = $\frac{402}{402+24} . 100\%$

= 94,36 %

Recall Positif = $\frac{1013}{1013+24} . 100\%$

= 97,69 %

Recall Negatif = $\frac{402}{402+0} . 100\%$

= 100 %

Pada Gambar 4.12 merupakan tabel pembuktian hasil dari klasifikasi menggunakan rulebase dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi pertama urut.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	naivebayes	rulebase	hasil
1	negatif	negatif	True Negatif
2	negatif	negatif	True Negatif
3	positif	positif	True Positif
4	positif	positif	True Positif
5	positif	positif	True Positif
6	positif	positif	True Positif
7	positif	positif	True Positif
8	positif	positif	True Positif
9	positif	positif	True Positif
10	positif	positif	True Positif
11	positif	positif	True Positif
12	negatif	negatif	True Negatif
13	negatif	negatif	True Negatif
14	positif	positif	True Positif
15	negatif	negatif	True Negatif
16	positif	positif	True Positif
17	negatif	negatif	True Negatif
18	positif	positif	True Positif
19	positif	positif	True Positif
20	negatif	negatif	True Negatif
21	negatif	negatif	True Negatif
22	positif	positif	True Positif

Showing 1 to 23 of 1,439 entries

Gambar 4.12. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*

Proporsi Pertama Acak

	testings	Freq
1	False Negatif	26
2	False Positif	2
3	True Negatif	380
4	True Positif	1031

Gambar 4.13. *Confusion Matrix* Proporsi Pertama Acak

Gambar 4.13 Menunjukkan *Confusion Matrix* untuk pengambilan data secara acak pada proporsi pertama

Tabel 4.2. *Confusion Matrix* Proporsi Pertama Acak

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	1031	26	1057
Negatif	2	380	380
Jumlah Kolom	1033	428	1461

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi pertama acak pada tokoh publik Joko Widodo

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Akurasi = $\frac{1031+380}{1031+380+2+26}$. 100%

= 98.05 %

Presisi Positif = $\frac{1031}{1031+2}$. 100%

= 99.8 %

Presisi Negatif = $\frac{380}{380+26}$. 100%

= 93,59 %

Recall Positif = $\frac{1031}{1031+26}$. 100%

= 97,54 %

Recall Negatif = $\frac{380}{380+2}$. 100%

= 99,47 %

Pada Gambar 4.14 merupakan tabel pembuktian hasil dari klasifikasi menggunakan rulebase dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi pertama acak.

	naivebayes	rulebase	hasil
1	negatif	negatif	True Negatif
2	negatif	negatif	True Negatif
3	positif	positif	True Positif
4	negatif	negatif	True Negatif
5	positif	positif	True Positif
6	positif	positif	True Positif
7	positif	positif	True Positif
8	negatif	negatif	True Negatif
9	positif	positif	True Positif
10	positif	positif	True Positif
11	negatif	negatif	True Negatif
12	positif	positif	True Positif
13	positif	positif	True Positif
14	positif	positif	True Positif
15	positif	negatif	False Negatif
16	negatif	negatif	True Negatif
17	negatif	negatif	True Negatif
18	negatif	negatif	True Negatif
19	positif	positif	True Positif
20	positif	positif	True Positif
21	positif	positif	True Positif
22	positif	positif	True Positif

Showing 1 to 23 of 1,439 entries

Gambar 4.14. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. KH. Mak'ruf Amin

Pada sistem yang pertama yaitu untuk tokoh publik KH. Ma'ruf Amin data yang digunakan berjumlah 11.391 data. Setelah melalui proses penghapusan data netral maka data yang diperoleh oleh tokoh publik KH. Ma'ruf Amin adalah 6.177 data. Pada pembagian proporsi data yang pertama yaitu 75% data training dan 25% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4632 data dan data testing sebanyak 1544 data.

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,74 %, presisi positif sebesar 100%, presisi negatif sebesar 97,68%, *recall* positif sebesar 99,70%, *recall* negatif sebesar 100% .

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,67 %, presisi positif sebesar 100%, presisi negatif sebesar 97,12%, *recall* positif sebesar 99,55%, recall negatif sebesar 100% .

3. Prabowo Subianto

Pada sistem yang pertama yaitu untuk tokoh publik Prabowo Subianto data yang digunakan berjumlah 15.600 data. Setelah melalui proses penghapusan data netral maka data yang diperoleh oleh tokoh publik Prabowo Subianto adalah 7130 data. Pada pembagian proporsi data yang pertama yaitu 75% data training dan 25% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 5347 data dan data testing sebanyak 1782 data.

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,43 %, presisi positif sebesar 99,76%, presisi negatif sebesar 98,53%, *recall* positif sebesar 99,46%, *recall* negatif sebesar 99,36% .

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,15%, presisi positif sebesar 99,75%, presisi negatif sebesar 97,75%, *recall* positif sebesar 99,04%, recall negatif sebesar 99,42% .

4. Sandiaga Salahudin Uno

Pada sistem yang pertama yaitu untuk tokoh publik Sandiaga Salahudin Uno data yang digunakan berjumlah 11.687 data. Seltelah melalui proses penghapusan data netral maka data yang diperoleh oleh tokoh publik Sandiaga Salahudin Uno adalah 5521 data. Pada pembagian proporsi data yang pertama yaitu 75% data training dan 25% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4140 data dan data testing sebanyak 1380 data.

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,42%, presisi positif sebesar 99,81%, presisi negatif sebesar 97,8%, *recall* positif sebesar 99,45%, *recall* negatif sebesar 99,25% .

Untuk hasil dari confusion matrik pada proporsi pertama pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,56%, presisi positif sebesar 99,9%, presisi negatif sebesar 98,48%, *recall* positif sebesar 99,02%, recall negatif sebesar 99,69% .

4.5.2 Skenario Proporsi Kedua 80 % Training 20 % Testing

1. Joko Widodo

Pada pembagian proporsi data yang kedua yaitu 80% data training dan 20% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4604 data dan data testing sebanyak 1151 data.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Proporsi Kedua Urut

	testings	Freq
1	False Negatif	15
2	True Negatif	319
3	True Positif	817

Gambar 4.15. *Confusion Matrix* Proporsi Kedua Urut

Gambar 4.15 Menunjukkan *Confusion Matrix* untuk pengambilan data secara urut pada proporsi kedua

Tabel 4.3. *Confusion Matrix* Proporsi Kedua Urut

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	817	15	832
Negatif	0	319	319
Jumlah Kolom	817	334	1151

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi kedua urut pada tokoh publik Joko Widodo

Akurasi = $\frac{817+319}{817+319+0+15}$. 100%

= 98.69 %

Presisi Positif = $\frac{817}{817+0}$. 100%

= 100 %

Presisi Negatif = $\frac{319}{319+15}$. 100%

= 95,5 %

Recall Positif = $\frac{817}{817+15}$. 100%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

= 98,19 %

Recall Negatif = $\frac{319}{319+0}$. 100%

= 100 %

Pada Gambar 4.16 merupakan tabel pembuktian hasil dari klasifikasi menggunakan *rulebase* dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi kedua urut.

	naivebayes	rulebase	hasil
1	positif	positif	True Positif
2	negatif	negatif	True Negatif
3	positif	positif	True Positif
4	positif	positif	True Positif
5	negatif	negatif	True Negatif
6	positif	positif	True Positif
7	positif	positif	True Positif
8	negatif	negatif	True Negatif
9	positif	positif	True Positif
10	negatif	negatif	True Negatif
11	positif	positif	True Positif
12	positif	positif	True Positif
13	positif	positif	True Positif
14	positif	positif	True Positif
15	negatif	negatif	True Negatif
16	negatif	negatif	True Negatif
17	positif	positif	True Positif
18	positif	positif	True Positif
19	negatif	negatif	True Negatif
20	negatif	negatif	True Negatif
21	negatif	negatif	True Negatif
22	positif	positif	True Positif

Showing 1 to 23 of 1,151 entries

Gambar 4.16. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes* Proporsi Kedua Acak

	testings	Freq
1	False Negatif	27
2	False Positif	3
3	True Negatif	310
4	True Positif	812

Gambar 4.17. *Confusion Matrix* Proporsi Kedua Acak

Gambar 4.17 Menunjukkan *Confusion Matrix* untuk pengambilan data secara acak pada proporsi kedua

Tabel 4.4. *Confusion Matrix* Proporsi Kedua Acak

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	812	27	839
Negatif	3	310	313
Jumlah Kolom	815	337	1152

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi kedua acak pada tokoh publik Joko Widodo

Akurasi = $\frac{812+310}{812+310+3+27}$. 100%

= 97.39 %

Presisi Positif = $\frac{812}{812+3}$. 100%

= 99,63 %

Presisi Negatif = $\frac{310}{310+27}$. 100%

= 91,98 %

Recall Positif = $\frac{812}{812+27}$. 100%

= 96,78 %

Recall Negatif = $\frac{310}{310+3}$. 100%

= 99,04 %

Pada Gambar 4.18 merupakan tabel pembuktian hasil dari klasifikasi menggunakan *rulebase* dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi kedua acak.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	naivebayes	rulebase	hasil
1	positif	positif	True Positif
2	positif	positif	True Positif
3	negatif	negatif	True Negatif
4	negatif	negatif	True Negatif
5	positif	positif	True Positif
6	positif	positif	True Positif
7	negatif	negatif	True Negatif
8	positif	positif	True Positif
9	positif	positif	True Positif
10	positif	positif	True Positif
11	positif	positif	True Positif
12	negatif	negatif	True Negatif
13	positif	positif	True Positif
14	positif	positif	True Positif
15	positif	positif	True Positif
16	negatif	negatif	True Negatif
17	positif	positif	True Positif
18	positif	positif	True Positif
19	positif	positif	True Positif
20	positif	positif	True Positif
21	positif	positif	True Positif
22	negatif	negatif	True Negatif

Showing 1 to 23 of 1,152 entries

Gambar 4.18. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*

2. KH. Ma'ruf Amin

Pada pembagian proporsi data yang kedua yaitu 80% data training dan 20% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4941 data dan data testing sebanyak 1235 data.

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,67 %, presisi positif sebesar 100%, presisi negatif sebesar 96,99%, *recall* positif sebesar 99,63%, *recall* negatif sebesar 100% .

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,43%, presisi positif sebesar 99,9%, presisi negatif sebesar 96,62%, *recall* positif sebesar 99,43%, recall negatif sebesar 99,42% .

3. Prabowo Subianto

Pada pembagian proporsi data yang kedua yaitu 80% data training dan 20% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 5704 data dan data testing sebanyak 1426 data.

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,29 %, presisi positif sebesar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

99,70%, presisi negatif sebesar 98,22%, *recall* positif sebesar 99,32%, *recall* negatif sebesar 99,23% .

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,64%, presisi positif sebesar 99,69%, presisi negatif sebesar 99,53%, *recall* positif sebesar 99,79%, recall negatif sebesar 99,30% .

4. Sandiaga Salahudin Uno

Pada pembagian proporsi data yang kedua yaitu 80% data training dan 20% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4416 data dan data testing sebanyak 1104 data.

Untuk hasil dari *confusion matrik* pada proporsi kedua pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,36 %, presisi positif sebesar 99,77%, presisi negatif sebesar 99,78%, *recall* positif sebesar 99,43%, *recall* negatif sebesar 99,10% .

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,54%, presisi positif sebesar 99,64%, presisi negatif sebesar 99,23%, *recall* positif sebesar 99,76%, recall negatif sebesar 98,85% .

4.5.3 Skenario Proporsi Ketiga 90 % Training 10 % Testing

1. Joko Widodo

Pada pembagian proporsi data yang ketiga yaitu 90% data training dan 10% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 5179 data dan data testing sebanyak 575 data.

Proporsi Ketiga Urut

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	testings	Freq
1	False Negatif	9
2	True Negatif	154
3	True Positif	412

Gambar 4.19. *Confusion Matrix* Proporsi Ketiga Urut

Gambar 4.19 Menunjukkan *Confusion Matrix* untuk pengambilan data secara urut pada proporsi ketiga

Tabel 4.5. *Confusion Matrix* Proporsi Ketiga Urut

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	412	9	421
Negatif	0	154	154
Jumlah Kolom	412	163	575

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi ketiga urut pada tokoh publik Joko Widodo

Akurasi = $\frac{412+154}{412+154+0+9}$. 100%

= 98.43 %

Presisi Positif = $\frac{412}{412+0}$. 100%

= 100 %

Presisi Negatif = $\frac{154}{154+9}$. 100%

= 94,47 %

Recall Positif = $\frac{412}{412+9}$. 100%

= 97,86 %

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Recall Negatif = $\frac{154}{154+0}$. 100%

= 100 %

Pada Gambar 4.20 merupakan tabel pembuktian hasil dari klasifikasi menggunakan Rulebase dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi ketiga urut.

	naivebayes	rulebase	hasil
1	positif	positif	True Positif
2	positif	positif	True Positif
3	positif	positif	True Positif
4	positif	positif	True Positif
5	positif	positif	True Positif
6	negatif	negatif	True Negatif
7	negatif	negatif	True Negatif
8	positif	positif	True Positif
9	positif	positif	True Positif
10	positif	positif	True Positif
11	negatif	negatif	True Negatif
12	negatif	negatif	True Negatif
13	positif	positif	True Positif
14	positif	positif	True Positif
15	positif	positif	True Positif
16	positif	positif	True Positif
17	positif	positif	True Positif
18	positif	positif	True Positif
19	positif	positif	True Positif
20	negatif	negatif	True Negatif
21	negatif	negatif	True Negatif
22	positif	positif	True Positif

Showing 1 to 23 of 575 entries

Gambar 4.20. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes* Proporsi Ketiga Acak

	testings	Freq
1	True Negatif	167
2	True Positif	409

Gambar 4.21. *Confusion Matrix* Proporsi Ketiga Acak

Gambar 4.21 Menunjukkan *Confusion Matrix* untuk pengambilan data secara acak pada proporsi ketiga

Tabel 4.6. *Confusion Matrix* Proporsi Ketiga Acak

Prediksi / Akurasi	Positif	Negatif	Total Baris
Positif	409	0	409
Negatif	0	167	167

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Jumlah Kolom	409	167	576

Berikut ini merupakan hasil perhitungan dari akurasi, presisi, dan recall dari pembagian proporsi ketiga acak pada tokoh publik Joko Widodo

Akurasi = $\frac{409+167}{409+167+0+0}$. 100%

= 100 %

Presisi Positif = $\frac{409}{409+0}$. 100%

= 100 %

Presisi Negatif = $\frac{167}{167+0}$. 100%

= 100 %

Recall Positif = $\frac{409}{409+0}$. 100%

= 100 %

Recall Negatif = $\frac{167}{167+0}$. 100%

= 100 %

Pada Gambar 4.22 merupakan tabel pembuktian hasil dari klasifikasi menggunakan *rulebase* dan hasil klasifikasi memakai metode *Naïve Bayes Classifier* memakai proporsi ketiga acak.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	naivebayes	rulebase	hasil
1	positif	negatif	False Negatif
2	positif	positif	True Positif
3	positif	positif	True Positif
4	positif	positif	True Positif
5	negatif	negatif	True Negatif
6	positif	positif	True Positif
7	negatif	negatif	True Negatif
8	negatif	negatif	True Negatif
9	positif	positif	True Positif
10	positif	positif	True Positif
11	positif	positif	True Positif
12	positif	positif	True Positif
13	positif	positif	True Positif
14	positif	positif	True Positif
15	positif	positif	True Positif
16	positif	positif	True Positif
17	positif	positif	True Positif
18	positif	positif	True Positif
19	positif	positif	True Positif
20	positif	positif	True Positif
21	positif	positif	True Positif
22	positif	positif	True Positif

Showing 1 to 23 of 576 entries

Gambar 4.22. Perbandingan *Confusion Matrix* Antara *Rulebase* dan *Naïve Bayes*

2. KH.Ma'ruf Amin

Pada pembagian proporsi data yang ketiga yaitu 90% data training dan 10% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 5559 data dan data testing sebanyak 617 data.

Untuk hasil dari confusion matrik pada proporsi ketiga pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,83 %, presisi positif sebesar 100%, presisi negatif sebesar 97,67%, *recall* positif sebesar 99,82%, *recall* negatif sebesar 100% .

Untuk hasil dari confusion matrik pada proporsi ketiga pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,51%, presisi positif sebesar 100%, presisi negatif sebesar 96,66%, *recall* positif sebesar 99,43%, recall negatif sebesar 100% .

3. Prabowo Subianto

Pada pembagian proporsi data yang ketiga yaitu 90% data training dan 10% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 6417 data dan data testing sebanyak 713 data.

Untuk hasil dari confusion matrik pada proporsi ketiga pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,43%, presisi positif sebesar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

99,79%, presisi negatif sebesar 98,61%, *recall* positif sebesar 99,39%, *recall* negatif sebesar 99,53% .

Untuk hasil dari confusion matrik pada proporsi ketiga pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,85%, presisi positif sebesar 100%, presisi negatif sebesar 99,46%, *recall* positif sebesar 99,81%, recall negatif sebesar 100% .

4. Sandiaga Salahudin Uno

Pada pembagian proporsi data yang ketiga yaitu 90% data training dan 10% data testing. Data yang diambil secara urut maupun acak terdiri dari data training sebanyak 4968 data dan data testing sebanyak 552 data.

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara urut menghasilkan nilai akurasi sebesar 99,45 %, presisi positif sebesar 99,76%, presisi negatif sebesar 98,33%, *recall* positif sebesar 99,53%, *recall* negatif sebesar 99,15% .

Untuk hasil dari confusion matrik pada proporsi kedua pengambilan data secara acak menghasilkan nilai akurasi sebesar 99,63%, presisi positif sebesar 99,72%, presisi negatif sebesar 100%, *recall* positif sebesar 100%, recall negatif sebesar 99,11% .

4.6. Hasil Analisis Rata-Rata Akurasi Pada Setiap Proporsi

1. Proporsi Pertama 75% Training dan 25% Testing

Pada bab ini akan dibahas mengenai presentasi rata rata nilai akurasi dari proporsi pertama baik pengambilan data secara urut maupun acak yang dimiliki oleh setiap tokoh publik tersebut. Dijelaksan dalam Tabel 4.7.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.7 Presentasi Nilai Akurasi Pada Proporsi Pertama

Tabel 4.7 Proporsi Pertama 75% Training 25% Testing

No	Nama	Data		Hasil Akurasi	
		Training	Testing	Urut	Acak
1	Joko Widodo	4316	1439	98,33%	98,05%
2	KH. Ma'ruf Amin	4632	1544	99,74%	99,67%
3	Prabowo Subiyanto	5347	1782	99,43%	99,15%
4	Sandiaga S Uno	4140	1380	99,42%	99,56%
	Total Rata-rata			**99,23%**	**99,11%**

Dalam tabel diatas dapat diketahui bahwa hasil rata rata akurasi pada proporsi pertama oleh keempat tokoh tersebut yaitu pada pengambilan data urut diperoleh rata rata akurasi sebesar 99,23% sedangkan dari pengambilan data secara acak pada proporsi pertama diperoleh rata rata akurasi sebesar 99,11%

2. Proporsi Pertama 80% Training dan 20% Testing

Pada bab ini akan dibahas mengenai presentasi rata rata dari nilai akurasi proporsi kedua baik pengambilan data secara urut maupun acak yang dimiliki oleh setiap tokoh publik tersebut. Dijelaksan dalam Tabel 4.8.

Tabel 4.8 Presentasi Nilai Akurasi Pada Proporsi Kedua

Tabel 4.8 Proporsi Pertama 80% Training 20% Testing

No	Nama	Data		Hasil Akurasi	
		Training	Testing	Urut	Acak
1	Joko Widodo	4604	1151	98,69%	97,39%
2	KH. Ma'ruf Amin	4941	1235	99,67%	99,43%
3	Prabowo Subiyanto	5704	1426	99,29%	99,64%
4	Sandiaga S Uno	4416	1104	99,36%	99,54%
	Total Rata-rata			**99,25%**	**99,00%**

Dalam tabel diatas dapat diketahui bahwa hasil rata rata akurasi pada proporsi kedua oleh keempat tokoh tersebut yaitu pada pengambilan data urut diperoleh rata rata akurasi sebesar 99,25% sedangkan dari pengambilan data secara acak pada proporsi pertama diperoleh rata rata akurasi sebesar 99,00%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3. Proporsi Pertama 90% Training dan 10% Testing

Pada bab ini akan dibahas mengenai presentasi rata rata dari nilai akurasi proporsi ketiga baik pengambilan data secara urut maupun acak yang dimiliki oleh setiap tokoh publik tersebut. Dijelasksan dalam Tabel 4.9.

Tabel 4.9 Presentasi Nilai Akurasi Pada Proporsi Ketiga

Tabel 4.9 Proporsi Pertama 90% Training 10% Testing

No	Nama	Data		Hasil Akurasi	
		Training	Testing	Urut	Acak
1	Joko Widodo	5179	575	98,43%	100%
2	KH. Ma'ruf Amin	5559	617	99,83%	99,51%
3	Prabowo Subiyanto	6417	713	99,43%	99,85%
4	Sandiaga S Uno	4968	552	99,45%	99,63%
	Total Rata-rata			**99,29%**	**99,75%**

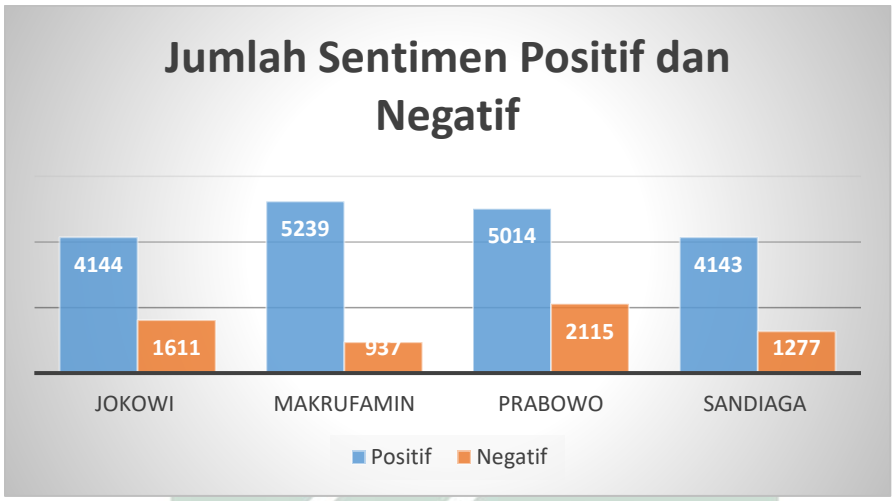
Dalam tabel diatas dapat diketahui bahwa hasil rata rata akurasi pada proporsi ketiga oleh keempat tokoh tersebut yaitu pada pengambilan data urut diperoleh rata rata akurasi sebesar 99,29% sedangkan dari pengambilan data secara acak pada proporsi pertama diperoleh rata rata akurasi sebesar 99,75%

Dari ketiga percobaan yang dilakukan maka dapat diambil kesimpulan jika metode *Naïve Bayes Classifier* ini cocok diterapkan pada proporsi ketiga dengan pengambilan data secara acak. Hasil rata rata yang diperoleh adalah 99,75%, hasil rata rata ini merpakan hasil yang paling tinggi diperoleh dibandingkan dengan percobaan yang lain.

4.7. Analisis Prosentasi Terhadap Jumlah Sentimen Positif dan Negatif

Setelah keempat sistem yang di buat diketahui memiliki nilai presisi, recall yang baik. Maka langkah selanjutnya adalah melakukan percobaan menggunakan lebih banyak data , sehingga akan diketahui kecendrungan sentimen masyarakat terkait masing masing tokoh publik. Berikut ini pada Gambar 4.23 adalah jumlah sentimen positif dan negatif sistem terhadap masing masing data:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.23. Tabel Jumlah Sentimen Positif dan Negatif

Pada Gambar 4.23, dapat dilihat data yang dipakai dalam penelitian kali ini berjumlah 24580 data yang terbagi ke dalam empat data yaitu data pertama untuk tokoh publik Joko Widodo yang kedua untuk tokoh publik KH. Ma'aruf Amin, yang ketiga untuk Prabowo Subianto, dan yang keempat untuk Sandiga Salahudin Uno. Data untuk tokoh publik yang pertama berjumlah 5755 data. Data untuk tokoh publik yang kedua berjumlah 6176 data. Data untuk tokoh publik yang ketiga berjumlah 7129 data. Data untuk tokoh publik yang keempat berjumlah 5511 data.

Untuk tokoh publik yang pertama yaitu Joko Widodo, sistem yang sudah dibuat dapat mengklasifikasiikan semua data yang ada menjadi 4144 data memiliki sentimen positif dan 1611 data yang memiliki sentimen negatif. Ini berartii dari data yang tersedia, tokoh publik yang pertama memperoleh sentimen positif sebanyak

$$\frac{4144}{5755} . 100\% = 72{,}01\ \%$$

Dan memperoleh sentimen negatif sebanyak

$$\frac{1611}{5755} . 100\% = 27{,}99\ \%$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dan untuk tokoh publik yang kedua yaitu KH. Ma'ruf Amin, sistem yang sudah dibuat mengklasifikasikan dapat semua data yang ada menjadi 5239 data memiliki sentimen positif dan 937 data memiliki sentimen negatif. Hal ini dikarenakan dari data yang ada, tokoh publik yang kedua memperoleh sentimen positif sebanyak

$$\frac{5239}{6176} . 100\% = 84{,}82\ \%$$

Dan memperoleh sentimen negatif sebanyak

$$\frac{937}{6176} . 100\% = 15{,}17\ \%$$

Untuk tokoh publik yang ketiga yaitu Prabowo Subianto, sistem yang sudah dibuat mengklasifikasikan semua data yang ada menjadi 5014 data memiliki sentimen positif dan 2115 data memiliki sentimen negatif. Hal ini dikarenakan dari data yang ada, tokoh publik yang kedua memperoleh sentimen positif sebanyak

$$\frac{5014}{7129} . 100\% = 70{,}33\ \%$$

Dan memperoleh sentimen negatif sebanyak

$$\frac{2115}{7129} . 100\% = 29{,}67\ \%$$

Untuk tokoh publik yang ketiga yaitu Sandiaga Salahudin Uno, sistem yang sudah dibuat mengklasifikasikan semua data yang ada menjadi 4243 data yang memiliki sentimen positif dan 1277 data memiliki sentimen negatif. Hal ini dikarenakan bahwa dari data yang ada, tokoh publik yang kedua mendapatkan sentimen positif sebanyak

$$\frac{4243}{5511} . 100\% = 76{,}99\ \%$$

Dan mendapatkan sentimen negatif sebanyak

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$\frac{1277}{5511} . 100\% = 23{,}17\ \%$$

4.8. Perhitungan Menggunakan *Metode Simple Weighted Product Model*

Pada bab ini membahas tentang perhitungan hasil dari proses klasifikasi topik dari sentimen twitter pada masing-masing tokoh publik. Pada proses perangkingan ini tidak bisa digunakan cara penambahan atau penjumlahan sentimen positif terbanyak saja, tetapi juga mempertimbangkan adanya sentimen nagatif yang ada pada setiap klasifikasi masing masing tokoh publik. Oleh sebab itu diperlukan suatu metode yang mempertimbangkan keduanya sebagai hasil dari proses perangkingan tersebut. Pada bab ini akan menghasilkan nilai rangking tiap-tiap tokoh publik dengan nilai sentimen yang didapatkan dengan menggunakan metode *Metode Simple Weighted Product Model.* Informasi sentimen tokoh politik diperoleh dari hasil klasifikasi tweet tokoh publik. Penggunaan metode *Metode Simple Weighted Product Model* dalam penentuan peringkat tokoh publik.

Ada 4 tokoh publik yang akan menjadi alternatif, yaitu:

Tabel 4.10 Tabel Alternatif Perhitungan SWP

Alternatif
Joko Widodo
KH.Ma'ruf Amin
Prabowo Subianto
Sandiaga Salahudin Uno

Tabel 4.11. Tabel Kriteria Perhitungan SWP

Kriteria
Positif
Negatif

Pengambil keputusan memberikan bobot preferensi sebagai berikut:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.12. Nilai Kriteria Setiap Alternatif Perhitungan SWP

No	Alternatif	Kriteria	
		Positif	Negatif
V1	Joko Widodo	4144	1611
V2	KH.Mak'ruf Amin	5239	937
V3	Prabowo Subianto	5014	2115
V4	Sandiaga Salahudin Uno	4243	1277
	Bobot	10	10

a. Proses perhitungan menggunakan metode Simple Weightes Product (SWP) akan dilakukan normalisasi bobot.

$$W_J = \frac{wj}{\sum wj}$$

W_1 = 10 / (10+10) = 10 / 20 = 0,5

W_2 = 10 / (10+10) = 10 / 20 = 0,5

b. Menghitung vector S

$$S_i = \prod_{J=1}^{n} X_{ij}^{\ w_j}$$

$S_1 = (4144^{0,5})\ (1611^{-0,5}) = 1,603844$

$S_2 = (5239^{0,5})\ (937^{-0,5}) = 2,364582$

$S_3 = (5014^{0,5})\ (2115^{-0,5}) = 2,539703$

$S_4 = (4243^{0,5})\ (1277^{-0,5}) = 1,822809$

c. Menghitung Vektor Perangkingan

$$V_i = \frac{\prod_{j=1}^{n} x_{ij}\ wj}{\prod_{j=1}^{n} (x_j *)wj}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

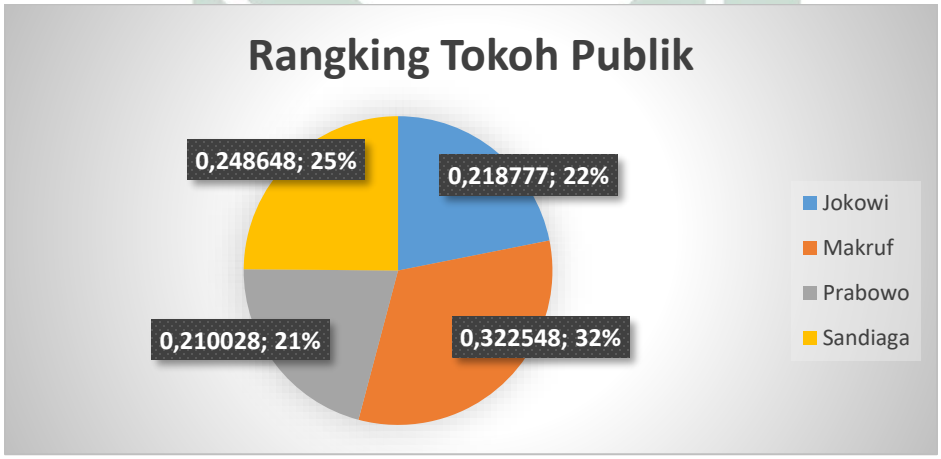
V_1 = 1,603844 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,218777

V_2 = 2,364582 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,322548

V_3 = 2,539703 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,210028

V_4 = 1,822809 / (1,603844 + 2,364582 + 2,539703 + 1,822809) = 0,248648

Dari proses perhitungan nilai sentimen dengan bobot topik diatas, didapatkan nilai akhir sentimen pada masing – masing tokoh publik dengan nilai tertinggi yakni 0,322548 (KH. Maruf Amin), 0,248648 (Sandiaga Salahudin Uno), 0,218777 (Joko Widodo) dan nilai terendah yakni 0,210028 (Prabowo Subianto). Maka kesimpulan didapatkan bahwa KH. Maruf Amin memiliki nilai citra positif tertinggi, disusul oleh Sandiaga Salahudin Uno kemudian Joko Widodo, dan yang terakhir yakni Prabowo Subianto. Agar lebih memudahkan untuk memahami akan disediakan sebuah diagram untuk hasil dari proses perangkingan yang sudah dilakukan.



Gambar 4.24. Diagram Rangking Tokoh Publik

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.9. Perbandingan Hasil Klasifikasi Antara Pelabelan Data Training Manual dan Rulebase Pada Proses *Naïve Bayes Classifier*

Pada bab kali ini akan dijelaskan mengenai percobaan antara pelabelan data training secara manual dan pelabelan data training menggunakan *rulebase* menggunakan metode *Naïve Bayes Classifier*. Percobaan ini dilakukan untuk mengetahui sejauh mana perbedaan nilai akurasi dari kedua macam pelabelan data training tersebut. Dengan menggunakan jenis data dan jumlah data yang sama yaitu 500 twit dari data milik Bapak Joko Widodo. Pada saat pelabelan data secara manual, didapatkan data netral sebanyak 116, data netral tersebut kemudian dihapus mengingat pelabelan dalam penelitian ini hanya sebatas sentiment positif dan negatif saja, sehingga menyisakan jumlah 384 yang terdiri dari data yang bersentimen positif dan negatif.

Sedangkan pada pelabelan data menggunakan rulebase data netral yang didapatkan sebanyak 392, data netral tersebut kemudian dihapus mengingat pelabelan dalam penelitian ini hanya sebatas sentiment positif dan negatif saja, sehingga menyisakan jumlah 108 yang terdiri dari data yang bersentimen positif dan negatif. Pada percobaan ini, peneliti melakukan percobaan dengan menggunakan skenario proporsi yang pertama (75% data training dan 25% data testing). Sehingga menghasilkan data berikut ini:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.9.1 Skenario Proporsi Pertama 75 % Training 25 % Testing

1. Manual

Tabel 4.13. Proporsi Pertama Urut Pada Pelabelan Secara Manual

Proporsi pertama Urut

NO	KETERANGAN	RINCIAN/CARA	HASIL
1	Akurasi	$\frac{59+17}{59+17+7+12} \cdot 100\%$	80%
2	Presi Positif	$\frac{59}{59+7} \cdot 100\%$	89,39%
3	Presi Negatif	$\frac{17}{17+12} \cdot 100\%$	58,62%
4	Recall Positif	$\frac{59}{59+12} \cdot 100\%$	83,10%
5	Recall Negatif	$\frac{17}{17+7} \cdot 100\%$	70,83%

Tabel 4.14. Proporsi Pertama Acak Pada Pelabelan Secara Manual

Proporsi pertama Acak

NO	KETERANGAN	RINCIAN/CARA	HASIL
1	Akurasi	$\frac{55+24}{55+24+4+13} \cdot 100\%$	82,29%
2	Presi Positif	$\frac{55}{55+4} \cdot 100\%$	93,22%
3	Presi Negatif	$\frac{24}{24+13} \cdot 100\%$	64,86%
4	Recall Positif	$\frac{55}{55+13} \cdot 100\%$	80,88%
5	Recall Negatif	$\frac{24}{24+4} \cdot 100\%$	85,71%

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Rulebase

Tabel 4.15. Proporsi Pertama Urut Pada Pelabelan Secara *Rulebase*

Proporsi pertama Urut			
NO	KETERANGAN	RINCIAN/CARA	HASIL
1	Akurasi	$\frac{62+7}{62+7+0+3} . 100\%$	95,83%
2	Presi Positif	$\frac{62}{62+0} . 100\%$	100%
3	Presi Negatif	$\frac{7}{7+3} . 100\%$	70%
4	Recall Positif	$\frac{62}{62+3} . 100\%$	95,38%
5	Recall Negatif	$\frac{7}{7+0} . 200\%$	100%

Tabel 4.16. Proporsi Pertama Acak Pada Pelabelan Secara Rulebase

Proporsi pertama Acak			
NO	KETERANGAN	RINCIAN/CARA	HASIL
1	Akurasi	$\frac{57+13}{57+13+0+3} . 100\%$	95,89%
2	Presi Positif	$\frac{57}{57+0} . 100\%$	100%
3	Presi Negatif	$\frac{13}{13+3} . 100\%$	81,25%
4	Recall Positif	$\frac{57}{57+3} . 100\%$	95%
5	Recall Negatif	$\frac{13}{13+0} . 100\%$	100%

Dalam percobaan kali ini dapat disimpulkan bahwa perbedaan perlakuan dalam pelabelan dapat mempengaruhi tingkat akurasi yang dihasilkan oleh sistem. Pada saat menggunakan pelabelan data training secara manual maka hasil akurasi yang didapatkan yaitu < 85 %. Sedangkan pelabelan data menggunakan rulebase menghasilkan nilai akurasi > 95%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pembuatan sistem klasifikasi dengan menggunakan *Naïve Bayes Classifier* dan menggunakan metode perhitungan *Simple Weighted Product Model* terhadap semua data yang diperoleh dari twitter, maka dapat disimpulkan bahwa:

1. Hasil yang di dapatkan menunjukkan bahwa metode *Naïve Bayes Classifier* cocok diterapkan di ketiga proporsi yang ada, karena sama sama mendapatkan nilai akurasi yang cukup tinggi. Tetapi setelah dilakukan rata rata maka percobaan yang paling sesuai yaitu menggunakan proporsi ketiga dengan pengambilan data secara acak dengan memperoleh hasil rata rata sebesar 99,75%.
2. Dengan menggunakan metode perhitungan Simple Weighted Product Model maka hasil yang didapatkan yaitu tokoh publik yang mendapatkan urutan (rangking) dalam sentimen masyarakat terkait citra positif yang dimiliki tokoh publik yang pertama adalah KH. Ma'ruf Amin dengan mendapatkan nilai 0,322548, setelah itu adalah Sandiaga Salahudin Uno yang mendapatkan nilai 0,248648, lalu Joko Widodo yang mendapatkan nilai 0,218777 dan nilai terendah yakni Prabowo Subianto dengan nilai 0,210028.
3. Perbedaan perlakuan dalam pelabelan dapat mempengaruhi tingkat akurasi yang dihasilkan oleh sistem. Pada saat menggunakan pelabelan data training secara manual maka hasil akurasi yang didapatkan yaitu < 85 %. Sedangkan pelabelan data menggunakan rulebase menghasilkan nilai akurasi > 95%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.2. Saran

Berdasarkan hasil penelitian dari analisis sentimen teks pada media sosial twitter menggunakan *Naïve Bayes Classifier* dan dengan menggunakan metode perhitungan *Simple Weighted Product Model* yang telah dilakukan, diperlukan penelitian yang lebih lanjut terkait dengan analisis sentimen tokoh publik ini. Berikut saran yang dapat disampaikan dari hasil kesimpulan diatas:

1. Pada Proses klasifikasi menggunakan *Naive Bayes* sangat bergantung dari data latih yang digunakan. Penggunaan data latih sebaiknya menggunakan proporsi data yang sama dari setiap kelasnya serta memastikan kelas dari data latih tersebut benar.
2. Pada proses crawling data twitter, data yang didapatkan dari masing masing tokoh publik banyak terdapat data outlier (data yang tidak sesuai). Oleh sebab itu crawling data sebaiknya dilakukan menggunakan cara lain atau dengan mengganti kata kunci yang lebih sesuai sehingga tidak didapatkan banyak data outlier
3. Untuk mendapatkan hasil yang signifikan, perlu dilakukan eksperimen penggunaan leksikon polaritas dan multi kata, menggunakan strategi yang sangat mendasar berdasarkan pencarian lemmas polaritas dalam teks / tweet. Jika tweet tidak mengandung setidaknya satu lemma juga ditemukan dalam leksikon polaritas eksternal, maka tweet tersebut tidak memiliki polaritas apa pun
4. Hasil akurasi yang didapatkan dalam penelitian kali ini belum tentu sama tingginya dengan hasil akurasi jika menggunakan metode yang berbeda. Hal ini bisa menjadi rekomendasi untuk penelitian selanjutnya untuk melakukan penelitian dengan Teknik yang sama tetapi menggunakan metode yang berbeda.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Daftar Pustaka

Asosiasi Penyelenggara Jasa Internet Indonesia – APJII, "Penetrasi & Perilaku Pengguna Internet Indonesia – survey 2019," p. 34, 2019.

Abdillah, S. (2011). *Penyakit Stroke Dengan Klasifikasi Data Mining Pada RS Santa Maria Pemalang*.

Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *JURNAL SAINS DAN SENI ITS Vol. 4, No.2*.

Balagatabi, Z. N., & Balagatabi, H. N. (2012). *Comparison of Decision Tree and SVM Methods in Classification of Researcher ' s Cognitive Styles in Academic Environment*. *1*(January), 31–43.

Borcard, D., Gillet, F., Legendre, P., Borcard, D., Gillet, F., & Legendre, P. (2011). Unconstrained Ordination. In *Numerical Ecology with R*. https://doi.org/10.1007/978-1-4419-7976-6_5

Chen, L. F., Su, C. T., & Chen, K. H. (2011). An improved particle swarm optimization for feature selection. *Intelligent Data Analysis*, *16*(2), 167–182. https://doi.org/10.3233/IDA-2012-0517

Copestake, A. (2007). *Natural Language Processing Outline of today ' s lecture*. (October).

Crawley, M. J. (2007). The R Book. Chichester: John Wiley & Sons. In *The R Book*. https://doi.org/10.1002/9780470515075

Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). Applied math and machine learning basics. *Deep Learning*, 25–26.

Hearst, M. A., & Hall, S. (n.d.). *P99-1001*.

Juniawan, I. (2009). *Klasifikasi Dokumen Teks Berbahasa Indonesia Menggunakan Minor Component Analysis*.

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

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*. https://doi.org/10.1017/CBO9780511809071

Mitchell, T. M. (2009). Machine learning in ecosystem informatics and sustainability. *IJCAI International Joint Conference on Artificial*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Intelligence. https://doi.org/10.1007/978-3-540-75488-6_2

Prasetyo, E. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. In *Penerbit Andi*. https://doi.org/10.1017/CBO9781107415324.004

Rini, D. C., Farida, Y., & Puspitasari, D. (2016). Klasifikasi Menggunakan Metode Hybrid Bayessian-Neural Network (Studi Kasus: Identifikasi Virus Komputer). *Jurnal Matematika "MANTIK," 1*(2), 38. https://doi.org/10.15642/mantik.2016.1.2.38-43

Saraswati, N. W. S. (2013). Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis. *Seminar Nasional Sistem Informasi Indonesia*, 586–591. Retrieved from http://is.its.ac.id/pubs/oajis/index.php/file/download_file/512

Syahmia Gusriani, Kartina Diah Kusuma Wardhani, M. I. Z. (2016). Jurnal Aksara Komputer Terapan Analisis Sentimen Berdasarkan Komentar Publik Terhadap Toko Online Pada Media Sosial Facebook (Studi Kasus : Zalora dan BerryBenka). *Jurnal Aksara Komputer Terapan Politeknik Caltex Riau*.

SYAKURO, A. (UNIVERSITAS I. N. M. M. I. M. (2017). PADA MEDIA SOSIAL MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC) DENGAN SELEKSI FITUR INFORMATION GAIN (IG) HALAMAN JUDUL SKRIPSI Oleh : ABDAN SYAKURO. *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*.

Ui, F. (2016). *Analisis sentimen..., Yislam, FASILKOM UI, 2016*.

Witten, Frank, & Hall. (2011). Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). In *Complementary literature None*.

Zuhri, F. N., & Alamsyah, A. (2017). *Menggunakan Naive Bayes Classifier Di Forum Kaskus Public Sentiment Analysis of Smartfren Brand Using Naive Bayes Classifier on Kaskus Forum*. *4*(1), 242–251.

Zuur, A. F., Ieno, E. N., Meesters, E., Zuur, A. F., Ieno, E. N., & Meesters, E. (2009). Accessing Variables and Managing Subsets of Data. In *A Beginner's Guide to R*. https://doi.org/10.1007/978-0-387-93837-0_3

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id