

Analisis Sentimen Masyarakat pada Media Sosial Twitter Terhadap Partai Politik Peserta Pemilu 2019 Menggunakan *Naïve Bayes Classifier*

Aprillia Rizki Adiati¹, Anisa Herdiani, S.T., M.T.², Widi Astuti, S.T., M.Kom.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹aprileee@students.telkomuniversity.ac.id,

²anisaherdiani@telkomuniversity.ac.id, ³widiwdu@telkomuniversity.ac.id

Abstrak

Menjelang Pemilihan Umum 2019, banyak partai politik memanfaatkan media sosial untuk berkampanye dan meningkatkan popularitas. Salah satu media sosial yang banyak digunakan dalam media promosi partai politik adalah twitter. Selain itu, media sosial twitter juga dapat dijadikan tempat oleh masyarakat dalam memberikan opini terhadap partai terkait baik opini positif maupun opini negatif. Pada tugas akhir ini dibuat untuk menganalisis opini masyarakat terhadap partai politik peserta pemilu 2019 menggunakan metode *Naïve Bayes Classifier*. Berdasarkan sistem yang dibangun, didapatkan hasil sentimen positif sebesar 53,8% dan sentimen negatif 46,13% dengan rata-rata akurasi sebesar 78,03%.

Kata kunci: pemilihan umum, partai politik, analisis sentimen, Twitter, Naïve Bayes Classifier.

Abstract

Ahead of the 2019 general election, many political parties used social media to campaign and increase popularity. One of the social media that is widely used in the media promotion of political parties is Twitter. In addition, social media twitter can also be used as a place by the public in providing opinions to parties related to both positive and negative opinions. In this final project, it is made to analyze public opinion towards political parties participating in the 2019 elections using the *Naïve Bayes Classifier* method. Based on the system that was built, obtained an average accuracy of 78.03% and the results of positive sentiment of 53.8% and negative sentiment of 46.13%.

Keywords: election, political parties, sentiment analysis, Twitter, Naïve Bayes Classifier

1. Pendahuluan

1.1 Latar Belakang

Saat ini pertumbuhan media sosial mengalami perkembangan yang sangat pesat. Menjelang Pemilihan Umum 2019, banyak partai politik memanfaatkan media sosial untuk berkampanye dan meningkatkan popularitas mereka. Salah satu media sosial yang banyak digunakan dalam media promosi partai politik adalah twitter.

Twitter sendiri telah banyak dimanfaatkan dalam pemilihan umum di beberapa negara seperti Singapura, Jerman, dan Amerika. [1] Menjelang pemilihan umum 2019, banyak partai yang berlomba-lomba dalam mempromosikan partainya melalui media sosial terutama twitter. Selain itu, media sosial twitter juga dapat dijadikan tempat oleh masyarakat dalam memberikan opini terhadap partai terkait baik opini positif maupun opini negatif.

Pada penelitian ini penulis menganalisa sentimen masyarakat pada twitter untuk memberikan informasi mengenai citra partai politik peserta pemilu 2019 menurut pandangan masyarakat. Metode yang digunakan yaitu *Naïve Bayes Classifier*. Berdasarkan 3 penelitian sebelumnya menghasilkan akurasi yang berbeda-beda, penelitian pertama yaitu mengenai Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode *Naïve Bayes Classifier* oleh Brata Mas Pintoko dengan hasil akurasi mencapai 86,8% [2], penelitian kedua yaitu mengenai Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan *Naïve Bayesian Classification* oleh Sandi Fajar Rodiyansyah dan Edi Winarko dengan hasil akurasi mencapai 93,58% [3], dan penelitian ketiga yaitu Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik pada Twitter oleh Ahmad Fathan Hidayatullah didapatkan hasil berupa prosentase akurasi pengujian klasifikasi fitur *term frequency* mencapai 70,88% sedangkan fitur TF-IDF mencapai 69,30% [1]. Metode ini dipilih karena terbukti bahwa metode *Naïve Bayes Classifier* memiliki beberapa kelebihan antara lain sederhana, cepat, berakurasi tinggi dan memiliki performansi yang cukup baik untuk melakukan klasifikasi.

1.2 Topik dan Batasan

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya topik yang dibahas pada penelitian ini adalah sebagai berikut.

- Bagaimana tingkat akurasi metode *Naïve Bayes Classifier* dalam mengklasifikasikan sentimen masyarakat terhadap partai politik peserta pemilu 2019?
- Bagaimana mengidentifikasi sentimen masyarakat terhadap partai politik peserta pemilu 2019?

Adapun beberapa batasan masalah terhadap sistem yang dibangun dalam penelitian ini adalah sebagai berikut.

- a. Dataset yang digunakan hanya berbahasa Indonesia yang berasal dari tweet di media sosial twitter.
- b. Dataset yang diambil seputar pemilu 2019.
- c. Dataset yang diambil bukan berupa link, berita, pertanyaan, foto, video, dan potongan tweet yang tidak mengandung arti atau yang bersifat netral.
- d. Klasifikasi dataset yang digunakan berupa analisis sentimen positif dan negatif.
- e. Dataset yang diambil hanya membicarakan mengenai partai politik bukan personal atau pribadi.
- f. Dataset yang diambil adalah tweet yang diambil saat menjelang pemilu sampai dengan pengumuman hasil akhir dari pemilu 2019.
- g. Dataset tweet yang diambil berdasarkan mention, hastag, dan pencarian langsung partai tersebut.
- h. Dataset partai politik yang diambil hanya yang berskala nasional.

1.3 Tujuan

Berdasarkan topik yang sudah dijelaskan sebelumnya, tujuan yang dicapai pada penelitian ini adalah sebagai berikut.

- a. Dapat mengetahui tingkat akurasi metode Naïve Bayes Classifier dalam mengklasifikasikan sentimen masyarakat terhadap partai politik peserta pemilu 2019.
- b. Dapat mengidentifikasi sentimen masyarakat terhadap partai politik peserta pemilu 2019.

1.4 Sistematika Penulisan

Jurnal penelitian ini terdiri dari beberapa sub-bagian yang setiap bagian memiliki peranan untuk menjelaskan sesuatu. Bagian pertama dimulai dari bagian latar belakang, identifikasi masalah, tujuan serta sistematika penulisan. Bagian kedua terdapat studi terkait. Bagian ketiga terdapat metodologi yang menjelaskan rancangan dan sistem atau produk yang dihasilkan. Bagian keempat terdapat evaluasi berisi hasil dari pengujian dan analisis hasil pengujian dari penelitian. Serta, pada bagian terakhir terdapat kesimpulan yang diambil dari hasil pengujian analisis hasil pengujian sehingga tidak ada kesimpulan dari teori ataupun nalar semata.

2. Studi Terkait

2.1 Twitter

Twitter menjadi layanan penghubung antar setiap orang. Setiap orang bisa bercerita tentang apapun mengenai aktivitas dan cara pandangnya. Pengguna lain siapapun itu dapat menyimak informasi tersebut secara kontinu. Dalam hal sosial politik, twitter juga ampuh digunakan sebagai media mengeluarkan pendapat atau melakukan kampanye.

Pada perkembangannya, twitter nan simpel itu pun menjadi sarana ampuh penyebar informasi kepada semua orang baik yang dikenal maupun tidak, dan untuk beragam keperluan bisa digunakan sesuai dengan tujuan yang diinginkan. [4]

2.2 Crawling Data

Crawling data adalah suatu proses untuk mengambil atau mengunduh data dari twitter dengan bantuan *Application Programming Interface (API)* twitter baik berupa data user maupun data tweet. [5] Crawling data ini dilakukan untuk mengambil data dari twitter dimana data tersebut dibutuhkan untuk tugas akhir ini. Cara melakukan crawling data yaitu dengan membuat dengan memasukkan kata kunci untuk mencari tweet yang sesuai dengan kebutuhan yaitu melalui mention, hastag, serta pencarian langsung partai terkait. Misalnya mencari kata kunci "@PDI_Perjuangan" maka program akan mengambil tweet yang menyebut akun tersebut. Pencarian langsung partai misalnya mencari kata kunci "PDIP" maka program akan mengambil semua data tweet yang berhubungan langsung mengenai PDIP.

2.3 Analisis Sentimen

Analisis sentimen adalah jenis pengolahan bahasa alami untuk mengetahui perasaan masyarakat mengenai produk atau topik tertentu. [6] Banyak metode yang dapat digunakan untuk menganalisis sentimen dari media sosial twitter, dan banyak yang mengembangkan metode yang sudah ada menjadi metode-metode baru yang berakurasi tinggi. Perbandingan beberapa metode seperti metode Naïve Bayes, K-nearest Neighbor, dan gabungan K-Means dan LVQ dalam mengklasifikasikan kategori buku berbahasa Indonesia dengan data yang digunakan berjumlah 200 buku, 150 buku digunakan sebagai data latih, sedangkan 50 buku digunakan sebagai data uji. Dari hasil penelitian yang dilakukan, metode KNN memperoleh akurasi sebesar 96% kemudian Naïve Bayes sebesar 98%, lalu kombinasi K-Means dan LVQ menghasilkan akurasi sebesar 92,2%. Metode Naïve Bayes mendapatkan hasil akurasi yang tertinggi. [7]

2.4 Naïve Bayes Classifier

Metode *Naïve Bayes Classifier* merupakan metode yang digunakan untuk memprediksi probabilitas. [8] Cara kerja metode ini yaitu dengan memprediksi peluang terjadinya kejadian di masa depan berdasarkan data yang ada pada sebelumnya. Tingkat performa dari sistem klasifikasi yang dibuat menggunakan *naïve bayes classifier* bergantung pada data yang dimiliki dan data yang dipilih sebagai data training. Jika data yang dipilih sebagai data training bisa mewakili semua atau sebagian besar data yang dimiliki, maka sistem klasifikasi yang dibuat mempunyai performa yang bagus.

Berikut merupakan rumus perhitungan dari metode *Naïve Bayes Classifier*:

- Rumus menghitung *prior probability*: $P(c) = \frac{c}{N}$
Dimana C: kelas pada data latih
N: total kelas C pada data latih
- Rumus menghitung nilai *conditional probability*: $P(w|c) = \frac{\text{count}(w,c)+1}{\text{count}(c)+|V|}$
Dimana W: token dari kata di dalam data latih
C: kelas pada data latih
|V|: total seluruh token kata pada kalimat di dalam data latih
- Rumus perhitungan tweet untuk penentuan kelas $P(C|D) = P(c) \times P(w|c)$
Dimana C: kelas pada data latih
D: data uji ke-n [9]

2.5 Cross Validation

Cross Validation adalah metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua segmen: satu digunakan untuk mempelajari atau melatih model dan yang lainnya digunakan untuk memvalidasi model.

Pengujian data menggunakan metode *cross validation* dimana data dibagi menjadi sejumlah *nfold* yang diinginkan, kemudian proses testing dan training dilakukan sebanyak *n* kali. Sampai hasil akurasi keluar, hasil akurasi tersebut didapat dari rata-rata akurasi pada setiap iterasi. [10]

2.6 Evaluasi

Precision dan recall merupakan teknik yang dapat digunakan untuk menghitung nilai performansi dari sistem pemrosesan teks. Perhitungan precision dan recall membutuhkan empat komponen yaitu TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative). Berikut merupakan rumus performansi yang akan diuji.

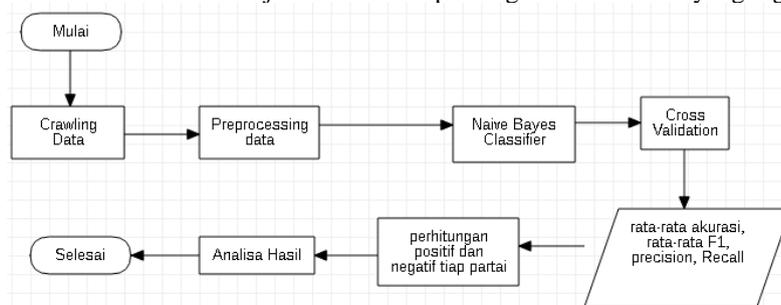
$$\text{Precision} = \frac{TP}{TP+FP} \quad (1) \qquad \text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (2) \qquad \text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

- Precision merupakan perbandingan atas jumlah klasifikasi yang benar dibandingkan dengan total item yang berhasil diklasifikasi di kelas positif.
- Recall** merupakan perbandingan atas jumlah klasifikasi yang benar dengan jumlah item yang seharusnya berada dalam kelas tersebut.
- F-measure** merupakan perhitungan rata-rata antara *precision* dan *recall*.
- Accuracy** merupakan perbandingan jumlah ketepatan klasifikasi atas semua item yang diklasifikasi. [11]

3. Metodologi

Sistem yang dibangun pada penelitian ini mengadaptasi dari jurnal untuk pengidentifikasian tweet mengenai partai politik peserta pemilu 2019 serta mendapatkan akurasi dari model yang digunakan. Data didapatkan dengan bantuan *Application Programming Interface (API)* yang disediakan oleh twitter. Setelah data terkumpul selanjutnya dilakukan *preprocessing* data untuk membersihkan data. Setelah data sudah bersih kemudian melakukan pemilihan data latih dan data uji. Berikut merupakan gambaran sistem yang digunakan.



Gambar 1 Tahapan proses dalam penelitian

Berdasarkan gambar diatas, penelitian ini diawali dengan *crawling* atau pengambilan data. Setelah pengambilan data selanjutnya adalah pembersihan data dengan proses *preprocessing* meliputi *cleansing*, *casefolding*, penghapusan *stopword*, dan tokenisasi,

3.1 Pengumpulan data

Pengumpulan data tweet yang dilakukan melalui API twitter. Pengumpulan data dilakukan dengan menggunakan library pada python yaitu Tweepy untuk mengakses twitter dari API itu sendiri. Setelah data terkumpul kemudian dilakukan pembersihan data dengan melihat tweet yang relevan mengenai partai politik peserta pemilu 2019. Data yang diambil berjumlah 1764 data dari total ada 16 partai yang diambil. Berikut merupakan contoh dataset yang telah dikumpulkan.

Tabel 0.1 pengumpulan data partai politik berdasarkan mention, hastag, dan pencarian langsung

No	Nama Partai	Singkatan Partai	Menggunakan @	Menggunakan #	Pencarian Langsung
1	Partai Kebangkitan Bangsa	PKB	@DPP_PKB	#PKB	PKB
2	Partai Gerindra	Gerindra	@Gerindra	#Gerindra	Gerindra
3	Partai Demokrasi Indonesia Perjuangan	PDIP	@PDI_Perjuangan	#PDIPerjuangan	PDIP
4	Partai Golongan Karya	Golkar	@PartaiGolkar	#Golkar	Golkar
5	Partai Nasional Demokrat	Nasdem	@NasDem	#Nasdem	Nasdem
6	Partai Garuda	Garuda	@DPPPpartaiGaruda	#PartaiGaruda	Garuda
7	Partai Berkarya	Berkarya	@berkaryaidofc	#PartaiBerkarya	Berkarya
8	Partai Keadilan Sejahtera	PKS	@PKSejahtera	#PartaiPKS	PKS
9	Partai Persatuan Indonesia	Perindo	@PartaiPerindo	#Perindo	Perindo
10	Partai Persatuan Pembangunan	PPP	@DPP_PPP	#PartaiPPP	PPP
11	Partai Solidaritas Indonesia	PSI	@psi_id	#PartaiPSI	PSI
12	Partai Amanat Nasional	PAN	@Official_PAN	#PartaiPAN	PAN
13	Partai Hati Nurani Masyarakat	Hanura	@PartaiHANURA	#Hanura	Hanura
14	Partai Demokrat	Demokrat	@PDemokrat	#Demokrat	Demokrat
15	Partai Keadilan dan Persatuan Indonesia	PKS	@PKSejahtera	#PKS	PKS
16	Partai Bulan Bintang	PBB	@PBB2019	#PartaiPBB	PBB

Tabel 0.2 Contoh Dataset

Tweet
Pada dasarnya mereka tidak suka kalo Presiden dr @PDIP...meski jelas2 Islam kalo dari @PDIP pasti dibilang kafir...Partai yang paling tidak bisa menerima kalo Presiden dari @PDIP adalah PKS...kalo Ormas dari FPI HTI dkk
Emang kader PDIP yang butuh nasi bungkus kwkwkw https://t.co/5Q7kDIW5oI

3.2 Pelabelan data

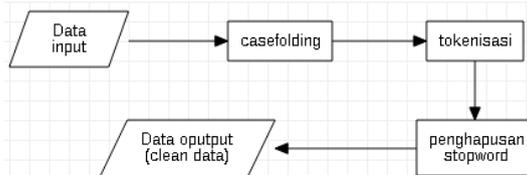
Proses pelabelan data dilakukan manual dengan menentukan suatu data masuk kedalam kelas positif atau kelas negatif. Proses pelabelan data dilakukan dengan melibatkan 3 validator untuk mengurangi penilaian secara subjektif. Pengelompokkan kelas positif dapat dilihat dari isi tweet yang mengandung kata positif, mendukung dan pernyataan setuju. Sedangkan kelas negatif dapat dilihat dari isi tweet yang mengandung kata bermakna negatif, ejekan, dan kontra.

Tabel 0.3 Contoh Pelabelan Data

Tweet	Label
indonesia maju bersama perindo	Positif
Heran PKPI tiap pemilu kecil pemilihnya tapi tiap pemilu ikutan mulu hehe	Negatif

3.3 Pre-processing data

Data yang didapatkan dari hasil pengumpulan data belum bisa langsung diklasifikasikan karena data tersebut masih terdapat banyak simbol dan kata-kata yang tidak diperlukan, karena itu diperlukan *preprocessing* agar data lebih terstruktur dan bersih sehingga bisa diklasifikasikan. Ada beberapa tahapan dalam *preprocessing* data diantaranya *casefolding*, penghapusan *stopword*, dan tokenisasi.



Gambar 2 Proses Preprocessing Data

Tahapan	Proses	Input	Output
Data Cleansing	Penghapusan website link, simbol, username, dan angka.	Pada dasarnya mereka tidak suka kalo Presiden dr @PDIP...meski jelas2 Islam kalo dari @PDIP pasti dibilang kafir...Partai yang paling tidak bisa menerima kalo Presiden dari @PDIP adalah PKS...kalo Ormas dari FPI HTI dkk	Pada dasarnya mereka tidak suka kalo presiden dari <u>PDIP</u> meski <u>jelas</u> Islam kalo dari <u>PDIP</u> pasti dibilang kafir partai yang paling tidak bisa menerima kalo Presiden dari <u>PDIP</u> adalah PKS kalo Ormas dari FPI HTI dkk
Casefolding	Perubahan dari huruf besar menjadi huruf kecil dan menghilangkan semua tanda baca pada kalimat.	Pada dasarnya mereka tidak suka kalo presiden dari PDIP meski jelas Islam kalo dari PDIP pasti dibilang kafir partai yang paling tidak bisa menerima kalo Presiden dari PDIP adalah PKS kalo Ormas dari FPI HTI dkk	<u>pada</u> dasarnya mereka tidak suka kalo presiden dari <u>pdip</u> meski jelas <u>islam</u> kalo dari <u>pdip</u> pasti dibilang kafir partai yang paling tidak bisa menerima kalo <u>presiden</u> dari <u>pdip</u> adalah <u>pk</u> s kalo ormas dari <u>fpi hti</u> dkk
Penghapusan Stopword	Penghapusan kata-kata yang bersifat umum namun tidak memiliki makna atau informasi yang dibutuhkan dan apabila dihapus tidak merubah makna pada tweet.	pada dasarnya mereka tidak suka kalo presiden dari pdip meski jelas islam kalo dari pdip pasti dibilang kafir partai yang paling tidak bisa menerima kalo presiden dari pdip adalah pks kalo ormas dari fpi hti dkk	pada dasarnya mereka tidak suka kalo presiden dari pdip meski jelas islam kalo dari pdip pasti dibilang kafir partai yang paling tidak bisa menerima kalo presiden dari pdip adalah pks kalo ormas dari fpi hti dkk
Tokenisasi	Pemisahan setiap kata pada tweet berdasarkan spasi yang ditemukan.	dasarnya mereka tidak suka presiden pdip meski jelas islam pdip pasti dibilang kafir partai paling tidak bisa menerima presiden pdip adalah pks ormas fpi hti dkk	Dasarnya; mereka; tidak; suka; presiden; pdip; meski; jelas; islam; pdip; pasti; dibilang; kafir; partai; paling; tidak; bisa; menerima; presiden; pdip; adalah; pks; ormas; fpi; hti; dkk;

3.4 Naïve Bayes Classifier

Pada implementasi *Naïve Bayes Classifier* digunakan data dari hasil *preprocessing*. Berikut merupakan contoh perhitungan probabilitas *prior*.

Tabel 0.5 Contoh Data Latih dan Data Uji

	Doc	Tweet	Kelas
Data Latih	1	Jangan samakan dengan kelakuan kalian yang berebut tulang dan recean	Negatif
	2	Haha partai baperan	Negatif
	3	Semoga tetap eksis	Positif
Data Uji	4	Dengan kelakuan eksis partai baperan	?

Langkah pertama yang dilakukan adalah menghitung nilai *prior probability* dari setiap kelas yang ada yaitu positif dan negatif.

$$P(\text{positif}) = \frac{1}{3} = 0,333 \qquad P(\text{negatif}) = \frac{2}{3} = 0,667$$

Setelah mendapatkan nilai *prior probability*, kemudian selanjutnya menghitung nilai *conditional probability*.

$$P(\text{dengan} | \text{negatif}) = \frac{1 + 1}{13 + 16} = \frac{2}{29} = 0,068 \qquad P(\text{dengan} | \text{positif}) = \frac{0 + 1}{3 + 16} = \frac{1}{19} = 0,052$$

$$P(\text{kelakuan} | \text{negatif}) = \frac{1 + 1}{13 + 16} = \frac{2}{29} = 0,068 \qquad P(\text{kelakuan} | \text{positif}) = \frac{0 + 1}{3 + 16} = \frac{1}{19} = 0,052$$

$$P(\text{eksis} | \text{positif}) = \frac{1 + 1}{3 + 16} = \frac{2}{19} = 0,105$$

$$P(\text{partai} | \text{positif}) = \frac{0 + 1}{3 + 16} = \frac{1}{19} = 0,052$$

$$P(\text{eksis} | \text{negatif}) = \frac{0 + 1}{13 + 16} = \frac{1}{29} = 0,034$$

$$P(\text{partai} | \text{negatif}) = \frac{1 + 1}{13 + 16} = \frac{2}{29} = 0,068$$

$$P(\text{baperan} | \text{negatif}) = \frac{1 + 1}{13 + 16} = \frac{2}{29} = 0,068$$

Setelah mendapatkan nilai *conditional probability*, selanjutnya dilakukan perhitungan untuk menentukan tweet tersebut masuk kedalam kelas positif atau kelas negatif. Berikut perhitungan tweet untuk penentuan kelas.

$$P(\text{negatif} | D4) = 0,333 \times 0,068 \times 0,068 \times 0,034 \times 0,068 \times 0,068 = 2,42079939 \times 10^{-7}$$

$$P(\text{positif} | D4) = 0,667 \times 0,052 \times 0,052 \times 0,105 \times 0,052 \times 0,052 = 5,12069027 \times 10^{-7}$$

Setelah didapatkan hasil probabilitas dari tweet terhadap setiap kelas, maka hasil probabilitas dibandingkan dari kedua kelasnya. Dilihat dari hasil diatas nilai probabilitas kelas positif lebih besar dibandingkan negatif sehingga D4 masuk kedalam kelas positif.

4. Evaluasi

Pengujian sistem pada penelitian ini dilakukan untuk melihat performansi dari metode *Naïve bayes classifier* yang dilakukan untuk melakukan analisis sentimen masyarakat terhadap partai politik peserta pemilu 2019 di twitter. Pengujian sistem menggunakan data sebanyak 1730 data tweet termasuk data uji dan data latih dimana data akan dibagi menjadi 2 bagian atau $k=2$ sehingga data yang diperoleh dibagi menjadi 2 bagian dengan masing masing 865 data. Data pada bagian pertama akan menjadi data uji dan data selanjutnya akan menjadi data latih. Selanjutnya akan seperti itu sampai dimana satu bagian untuk data uji akan selalu bergeser ke bagian kiri dan sisanya akan menjadi data latih. Sehingga terdapat hasil pengujian yaitu *precision*, *recall*, dan *accuracy* pada *Naïve Bayes Classifier*.

4.1 Hasil Pengujian

Berikut merupakan hasil dari pengujian menggunakan $k\text{-fold}=2$.

Tabel 0.6 Hasil Pengujian Model

Precision	Recall	Rata-rata F1-Measure	Rata-rata Accuracy
55,47%	66,13%	80,16%	78,03%

Dari hasil pengujian pada *Naïve Bayes Classifier*, didapatkan nilai rata-rata F1 measure 80,16% dan rata-rata Akurasi 78,03%. Berdasarkan hasil tersebut, juga didapatkan hasil prediksi positif dan negatif terhadap partai seperti berikut,



Gambar 3 Hasil Persentase Positif dan Negatif seluruh partai

4.2 Analisis Hasil Pengujian

Dari hasil pengujian diatas, ada beberapa faktor yang dapat mempengaruhi hasil performansi sistem diantaranya banyaknya data, perbandingan data antara sentimen positif dan negatif, dan juga nilai dari $k\text{-fold}$ nya. Berikut merupakan nilai rata-rata F1 dan rata-rata akurasi dari $k=2$ sampai dengan $k=10$.

Tabel 0.7 tabel pengujian berdasarkan K-Fold

K-Fold	Nilai rata-rata F1	Nilai rata-rata akurasi
2	0,80	0,78
3	0,70	0,65
4	0,68	0,64
5	0,62	0,59
6	0,61	0,59
7	0,54	0,54
8	0,56	0,57
9	0,52	0,55
10	0,48	0,51

Berdasarkan tabel diatas, nilai rata-rata akurasi dan F1 yang terbaik terletak pada k=2. sebuah himpunan secara acak menjadi 2 subset yang saling bebas. Dilakukan pengulangan sebanyak 2kali untuk pelatihan dan pengujian. Pada setiap ulangan, disisakan satu subset untuk pengujian dan subset lainnya untuk pelatihan. Tingkat akurasi dihitung dengan membagi jumlah keseluruhan klasifikasi yang benar dengan jumlah semua pada data awal. Semakin besar k-fold yang digunakan maka semakin besar perulangan yang dilakukan dan semakin kecil nilai rata-rata akurasi serta nilai F1 yang dihasilkan.

5. Kesimpulan

Berdasarkan hasil analisis terkait sentimen terhadap partai politik peserta pemilu 2019 dengan menggunakan *Naïve Bayes Classifier* untuk menentukan kelas prediksi, didapatkan hasil sentimen negatif sebesar 46,13% dan sentimen positif sebesar 53,87% dengan akurasi sebesar 78,03%.

Sedangkan untuk pengembangan penelitian selanjutnya, ada beberapa hal yang direkomendasikan yaitu sebagai berikut.

1. Dapat menangani pengambilan data diluar dari yang dibutuhkan seperti contoh PAN dan PBB tersebut.
2. Dapat menangani berbagai bahasa, tidak hanya Bahasa Indonesia saja namun meliputi Bahasa Inggris maupun bahasa daerah.
3. Menggunakan spesifikasi PC yang lebih tinggi lagi agar mendapatkan hasil akurasi yang lebih besar.
4. Pengambilan dataset sebaiknya dilakukan seminggu sekali agar mendapatkan variasi dataset yang lebih banyak lagi.

- [1] A. F. Hidayatullah, "Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter," 2014.
- [2] B. M. Pintoko, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," 2018.
- [3] S. F. R. d. E. Winarko, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," 2013.
- [4] i. maulana, "istilah-istilah dasar dalam twitter (pemula)," 2012.
- [5] E. B. S. Z. A. B. Jaka Eka Sembodo, "Data Crawling Otomatis pada Twitter," *indonesian symposium on computing (INDOSC)*, 2016.
- [6] G. V. d. R. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey," *International Journal of Advanced Research in Computer Science and Software Engineering*, 2012.
- [7] C. Darujati, "Pemanfaatan Teknik Supervised untuk Klasifikasi teks Bahasa Indonesia," *Jurnal Bandung Text Mining*, 2012.
- [8] F. d. P. Rini, "Klasifikasi Menggunakan Metode Hybrid Bayyesian-Neural Network (Studi Kasus: Identifikasi Virus Komputer)," *Jurnal Tematik "MANTIK"*, 2016.
- [9] R. S. P. M. A. F. Agnes Rossi T. L., "Analisis Sentimen Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2017.
- [10] P. Pitria, "Analisis Sentimen Pengguna Twitter pada Akun Resmi Samsung Indonesia dengan Menggunakan Naive Bayes," 2014.
- [11] E. Prastyo, "DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab," 2014.