



Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan *Naive Bayes*

Safitri Juanita

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: Safitri.juanita@budiluhur.ac.id

Abstrak—Menurut evaluasi BAWASLU berbagai konten negatif terkait dukung mendukung pasangan calon menyeruak ke berbagai laman media sosial. Tak pelak, terkadang konten tersebut menyerempet ke isu hoaks (kabar bohong) hingga ke isu suku ras agama dan antar golongan (SARA). Salah satu media sosial yang digunakan oleh masyarakat Indonesia adalah Twitter, menurut Kompas.com Jumlah pengguna harian Twitter secara global diklaim meningkat, ini terlihat ada laporan finansial Twitter kuartal ke-3 tahun 2019 pada laporan finansial Twitter kuartal ke-3 tahun 2019, pengguna aktif harian di platform Twitter dicatat meningkat 17 persen, ke angka 145 juta pengguna. Sehingga diperlukan sebuah penelitian analisis sentimen yang dapat menangkap pola persepsi masyarakat di media sosial twitter terhadap pemilu 2019 dan diharapkan penelitian ini dapat membantu pihak-pihak yang berkepentingan untuk meningkatkan Tingkat Partisipasi Pemilih pada 5 tahun kedepan. Metode penelitian ini menggunakan data tweet bahasa Indonesia yang diambil dari 16 April 2018-16 April 2019, selanjutnya data di preprocessing, text transformation, stemming bahasa Indonesia, penentuan class atribut, load dictionary dan dilakukan klasifikasi Naive Bayes menggunakan Weka. Kesimpulan dari penelitian ini adalah klasifikasi Naive Bayes menemukan bahwa dataset tweet pemilu 2019 memiliki pola persepsi negatif sebesar 52% jauh lebih besar dari persepsi positif yaitu 18% dan Persepsi netral memiliki nilai 31% lebih tinggi dari persepsi positif. Tingkat akurasi klasifikasi Naive Bayes terhadap dataset training bernilai 81% dan dataset testing 76%, rata-rata nilai precision untuk sentimen positif 86,65%, sentimen negatif 77,15%, dan sentimen netral bernilai 80,95% sedangkan rata-rata nilai recall pada sentimen positif 36,8%, sentimen negatif 93,2% dan sentimen netral 86,8%.

Kata Kunci: Persepsi Masyarakat, Pemilu 2019, Twitter, *Naive Bayes*, Weka

Abstract—According to the BAWASLU evaluation a variety of related negative content supports supporting prospective couples to burst into various social media pages. So sometimes the content leads to a hoax issue to the issue of religious and inter-group Racial (SARA). One of the social media used by the people of Indonesia is Twitter, according to Kompas.com number of Twitter daily users globally claimed to be increasing, this appears to be the 3rd Quarter Twitter Financial Report of 2019 on Twitter's 3rd quarter of 2019 Financial reports, daily active users on the Twitter platform are recorded to increase by 17 percent, to the number of 145 million users. So it is necessary that a sentiment analysis study can capture a pattern of community perception on social media Twitter against the 2019 elections and it is expected that this research can help interested parties to increase voter participation rate in the next 5 years. This research method uses the Indonesian tweet data taken from 16 April 2018-16 April 2019, further data in preprocessing, text transformation, stemming Bahasa Indonesia, specifying attribute class, load dictionary and a classification of Naive Bayes using Weka. The conclusion of this study was the classification of Naive Bayes finding that the 2019 election tweet dataset had a negative perception pattern of 52% much greater than the positive perception of 18% and the neutral perception had a value of 31% higher than positive perception. Naive Bayes ' degree of classification accuracy against the training dataset is 81% and the dataset testing 76%, the average precision value for positive sentiment is 86.65%, negative sentiment is 77.15%, and neutral sentiment is worth 80.95% while the average recall rate on positive sentiment is 36.8%, negative sentiment is 93.2% and the neutral sentiment is 86.8%.

Keywords: Community Perception, Elections 2019, Twitter, Naive Bayes, Weka

1. PENDAHULUAN

Menurut Henry B Mayo dikutip oleh Sorik [1] sebagai negara demokrasi, pelaksanaan pemilu salah satu syarat penting bagi terciptanya sebuah negara yang demokratis. Pesta Demokrasi di Indonesia diwujudkan dalam bentuk pemilihan umum (pemilu) presiden dan anggota legislatif yang dilakukan setiap 5 tahun sekali, namun berdasarkan data KPU, tingkat Partisipasi Pemilih (TPP) di Indonesia setelah era pasca reformasi mengalami penurunan setiap periode pemilu [2] dimana sebelum reformasi pada pemilu 1997, pemilu terakhir orde baru, tingkat partisipasi pemilih 93,6% namun pada pemilu legislatif tahun 2009, partisipasi pemilih menurun hingga 70,9% dan partisipasi pemilih Pilpres 2009 71,7%. Pada pemilu legislatif 2014, partisipasi pemilih mencapai 72%. Sedangkan pada Pilpres 2014, partisipasi pemilih mencapai 69,58% [3].

Pentingnya partisipasi masyarakat dalam Pemilu, sama pentingnya dengan upaya memperdalam proses demokrasi di tingkat masyarakat secara luas. Jika prasyarat standar demokrasi adalah terlaksananya Pemilu, maka partisipasi adalah salah satu indikator kualitas demokrasi tersebut [4] Salah satu alasan ketidakhadiran pemilih terutama di daerah perkotaan adalah karena informasi yang mereka dapatkan tentang politik justru membuat mereka menjadi apatis terhadap politik sehingga enggan berpartisipasi dalam pemilu, sedangkan rendahnya tingkat partisipasi pemilih kalangan muda atau pemilih muda disebabkan banyaknya komentar-komentar negatif mengenai politik dan politisi. Menurut Evaluasi Bawaslu [4] Berbagai konten negatif terkait dukung mendukung pasangan calon menyeruak ke berbagai laman media sosial. Tak pelak, terkadang konten tersebut menyerempet ke isu hoaks (kabar bohong) hingga ke isu suku ras agama dan antar golongan (SARA).

Salah satu media yang digunakan oleh masyarakat di Indonesia untuk mendapatkan informasi adalah media sosial, salah satunya adalah Twitter. Menurut Bill Clinton [5] Jumlah pengguna harian Twitter secara global



diklaim meningkat. Pada laporan finansial Twitter kuartal ke-3 tahun 2019, pengguna aktif harian di platform Twitter dicatat meningkat 17 persen, ke angka 145 juta pengguna. Menyikapi semakin rendahnya partisipasi masyarakat dalam pemilu maka penelitian ini ingin menangkap pola persepsi masyarakat di media sosial Twitter pada pemilu 2019 dengan menggunakan analisis sentimen dan klasifikasi Naive Bayes, diharapkan hasil dari penelitian ini dapat membantu pihak-pihak yang berkepentingan untuk meningkatkan tingkat partisipasi pemilih (TPP) pada 5 tahun kedepan. Penelitian analisis sentimen menurut Pan,dkk [6] selain sentimen positif dan negatif, ada juga sentimen netral dan sentimen campuran. Sentimen campuran berarti sentimen pengguna positif dalam beberapa aspek tetapi negatif pada yang lain. Sentimen Netral berarti tidak ada sentimen yang diungkapkan oleh pengguna.

Sentimen Analisis pada media sosial twitter sudah dilakukan oleh beberapa peneliti diantaranya yaitu penelitian Simorangkir dan Lhaksmana [7] yang melakukan penelitian sentimen analisis menggunakan metode Naive Bayes terhadap penggemar game online mobile Legends dan Arena Of Valor yang menyebabkan penggembar kedua permainan tersebut saling membanding-bandingkan dan menjadi perdebatan bagi para pencintanya, data tweet berbahasa inggris dikumpulkan dari tanggal 28 Mei 2018 – 17 Juli 2018. Penelitian ini menyimpulkan bahwa Naive Bayes Classification dapat memprediksi polarisasi sentiment Mobile Legends dengan nilai hasil akurasi, error, recall dan precision yang didapat masing-masing sebesar 88,89%, 19,18%, 96,97% dan 69,57% sedangkan Arena of Valor memiliki nilai akurasi, eror, recall dan precision masing-masing sebesar 39,02%, 60,98%, 88,89% dan 28,74%. Penelitian lain dilakukan Buntoro [8] yang melakukan analisis sentimen calon gubernur DKI Jakarta tahun 2017 menggunakan klasifikasi Naive Bayes dan Support Vector Machine, dataset menggunakan data Twitter berbahasa Indonesia dengan kata kunci AHY, AHOK, ANIES sebanyak 300 tweet. kesimpulan penelitian ini adalah Akurasi tertinggi didapat saat menggunakan metode klasifikasi Naive Bayes Classifier (NBC), dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%. penelitian lainnya dilakukan oleh Cahyo,dkk [9] yang meneliti sentimen masyarakat menggunakan data Twitter selama 2 pekan (15 Maret 2019-18 April 2019) terhadap pasangan calon presiden 2019. Analisis menggunakan sentiment analysis Tools NRC lexicon dan Bing Lexicon Tools, kesimpulan penelitian ini adalah untuk kandidat 01 Joko Widodo, analisis NRC lexicon memberikan nilai sebesar 249 dan bing lexicon sebesar 267 dengan nilai rata-rata 0,11, sedangkan untuk kandidat 02 Prabowo Subianto analisis NRC lexicon memberikan nilai sebesar 195 dan bing lexicon sebesar 204 dengan nilai rata-rata 0,085. Menggunakan data bulan april setelah pemilu 2019. Kandidat 01 Joko Widodo tetap mendapatkan respon yang banyak dari masyarakat netizen tetapi nilai sentimennya bergeser ke arah lebih negatif dibandingkan dengan kandidat 02 Prabowo Subianto. Untuk kandidat 01 Joko Widodo analisis NRC lexicon memberikan nilai sebesar 17 dan bing lexicon sebesar -273 dengan nilai rata-rata -0,246, sedangkan untuk kandidat 02 Prabowo Subianto analisis NRC lexicon memberikan nilai sebesar 238 dan bing lexicon sebesar -73 dengan nilai rata-rata -0.02430939. Penelitian yang mirip Cahyo, dkk namun dengan metode yang berbeda dilakukan oleh Lingga, dkk [10] penelitian ini melakukan sentimen analisis masyarakat di Twitter terhadap hasil *Quick Count* pada pemilihan presiden Indonesia 2019 dengan menggunakan metode Naive Bayes. kesimpulan penelitian ini adalah berdasarkan hasil pengolahan data menggunakan klasifikasi Naive Bayes menemukan bahwa tingkat akurasi sebesar 82,90% dengan $\alpha = 0,05$. Klasifikasi yang diperoleh masing-masing sebesar 34,5% (471) tweet positif dan 65,5% (895) tweet negatif terhadap hasil quick count. perbedaan penelitian yang sudah dilakukan sebelumnya terhadap penelitian ini adalah pada penelitian ini melakukan sentimen analisis persepsi masyarakat terhadap pemilu 2019 secara keseluruhan baik itu persepsi pemilihan presiden maupun pemilihan legislatif berbeda dengan penelitian [9][10] dan kesamaan penelitian sebelumnya dengan penelitian ini adalah kesamaan dalam mengambil dataset yaitu melalui media sosial Twitter seperti penelitian [7][8][9][10] dan menggunakan klasifikasi Naive bayes seperti penelitian [7][8][10].

2. METODOLOGI PENELITIAN

2.1 Desain Penelitian

Langkah-langkah yang akan dilakukan dalam penelitian terdapat pada gambar 1: (1) Melakukan studi literatur penelitian sebelumnya serta studi literatur definisi serta hal-hal yang berkaitan dengan text mining. (2) *Business Understanding*, analisa masalah yang saat ini sedang berkembang di masyarakat. (3) *Data Understanding*, pengumpulan tweet (*crawling tweet*) [11] menggunakan API Twitter dari tanggal 16 April 2018 – 16 April 2019. Pencarian opini menggunakan metode pengumpulan tweets manual yaitu memasukkan berbagai macam kata kunci yang berhubungan dengan topik pemilu 2019 pada kolom pencarian twitter dengan menggunakan kata pencarian : pemilu, pemilu 2019 ataupun pencarian dengan hastag contoh #pemilu #pemilu2019 serta data yang akan diambil berdasarkan mention yang masuk pada akun twitter resmi @KPU_ID. penentuan sentimen secara manual, pembagian dataset 70% data training dan 30% data testing. (4) *Data preprocessing Tweet (Case Folding, Cleasing, Tokenizer, Normalization*, Menghilangkan huruf yang berulang, Menghilangkan *emoticon*). (5) *Stemming*, yaitu mengambil kata dasar dalam tweet dengan menghilangkan awalan, akhiran atau keduanya. (6) Penentuan class attribute dan Load Dictionary yaitu mencocokkan kata dasar dengan kamus kata sentimen untuk mengetahui kandungan sentimen (positif, netral, negatif). Seluruh data tweet dilabelkan menurut kelas, kelas yang akan dipakai pada penelitian ini ada 3, yaitu kelas positif, kelas negatif dan kelas netral. Proses labelling tweet dilakukan secara



manual. Setelah data sudah diberikan label, data dibagi menjadi data training dan data testing menurut tanggal pengambilan tweet. (7) Klasifikasi data training dengan WEKA menggunakan klasifikasi Naive Bayes. Data training berupa *file* dengan format *.arff* yang berisi nilai dari atribut-atribut yang ditentukan sesuai dengan parameter yang digunakan (jumlah kata positif, jumlah kata netral, jumlah kata negatif). Hasil klasifikasi adalah kandungan sentimen (positif, netral, negatif) [6] pada setiap data dalam data training.



Gambar 1. Desain Penelitian

2.2 Algoritma Klasifikasi Naive Bayes

Algoritma klasifikasi Naive Bayes menurut [12] yang mengutip [13] merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat, pengklasifikasi Bayesian memiliki akurasi dan kecepatan yang lebih tinggi, terutama bila diterapkan pada dataset besar [12]. Gambar 2 adalah rumus Naive Bayes [13] :

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}. \quad (1)$$

Pada penelitian ini, nilai probabilitas dihitung dari kemunculan opini yang setara dengan perkalian nilai probabilitas kemunculan fitur dalam opini tersebut.

2.3 Perangkat Lunak Data Mining WEKA

WEKA *Waikato Environment for Knowledge Analysis* (WEKA) [14] dikembangkan di *University of Waikato*. WEKA adalah perangkat lunak *machine learning* tidak berbayar yang dibangun dalam bahasa JAVA. WEKA memiliki banyak data preprocessing dan teknik permodelan, mendukung beberapa task data mining standar, khususnya seperti data *preprocessing*, *clustering*, *classification*, *regression*, *visualization* dan *feature selection*. pada penelitian ini digunakan WEKA 3.9.

2.4 Evaluasi Klasifikasi Model

Menurut [12] mengutip [13] Untuk mengevaluasi model klasifikasi, dataset yang digunakan dibagi menjadi dua bagian: satu untuk pelatihan dan satu untuk pengujian. Biasanya, Bagian data latihan lebih dari bagian data pengujian. Tergantung pada ukuran data dan strategi evaluasi, rasio dari dua bagian mungkin berbeda. *Cross-Validation* adalah metode evaluasi yang banyak digunakan. Untuk mengevaluasi kinerja *classifier*, beberapa metrik yang digunakan untuk mengukur akurasi, presisi dan Recall dari prediksi atau rekomendasi.

3. HASIL DAN PEMBAHASAN

3.1 Mengumpulan Data Tweet

Proses *Crawling* Data menggunakan API Twitter dilakukan dari tanggal tanggal 16 April 2018 sampai tanggal 16 April 2019 dengan kata kunci “PEMILU”, “PEMILU 2019”, “PILKADA2019”, “PILKADA” Dari proses *crawling* data twitter didapatkan data sebanyak 221 data tweet yang belum diketahui label sentimennya.



3.2 Penentuan Sentimen secara manual

Tahap selanjutnya dari penelitian ini adalah menentukan label sentimen secara manual pada setiap data tweet, yang berjumlah 221 data. Penentuan label terdiri dari tiga kategori yaitu positif, negatif dan netral. Pada penelitian ini pemberian label sentimen dilakukan oleh tim ahli bidang ilmu komunikasi yang lebih memahami opini masyarakat dilihat dari sisi komunikasi yang pada penelitian ini menggunakan nara sumber Linda Islami, M.Si dan Dr. Nawiroh Vera, S.Sos, M.Si

3.3 Pembagian Dataset

Setelah dilakukan penentuan sentimen secara manual oleh tim ahli terhadap 221 data tweet, maka data tweet dikelompokkan menjadi beberapa sentimen pada tabel 1 berikut :

Tabel 1. Detail Dataset Tweet

Positif	Negatif	Netral	Total
58	101	62	221

Setelah diketahui sentimen pada setiap tweet maka langkah selanjutnya adalah memisahkan dataset dengan pembagian 70% data training dan 30% data testing. Sampel pemilihan dataset menggunakan teknik *stratified modeling sample* yaitu sebuah teknik sampel terstratifikasi dengan populasi dibagi atas kelompok-kelompok yang homogen. Dari 221 jumlah dataset akan dibagi menjadi 154 data tweet untuk data training, yang terdiri dari 41 sentimen positif, 70 sentimen negatif dan 43 sentimen netral. Sedangkan untuk data testing terdiri dari 67 data tweet terdiri dari 17 sentimen positif, 31 sentimen negatif dan 19 sentimen netral. Berdasarkan total keseluruhan pembagian data, maka hasil tersebut ditampilkan pada tabel 2 di bawah ini :

Tabel 2. Hasil Pembagian Data Training dan Data Testing

Sentimen	Data Training	Data Testing	Total
Positif	41	17	58
Negatif	70	31	101
Netral	43	19	62
Total	154	67	221

3.4 Hasil Data Preprocessing

Setelah dilakukan proses *case folding* (merubah menjadi huruf kecil), *cleasing (noise)*, kata yang dihilangkan adalah karakter HTML, titik, koma, tanda seru, tanda tanya, emoticon, hashtag (#), username, URL dan email), *tokenizing* (memisahkan kalimat menjadi sebuah kata tunggal), *normalization* (merubah kata-kata tidak yang disingkat menjadi kata yang bisa dipahami), *stopword removal* (membuang kata-kata yang tidak memiliki arti atau tidak relevan), *stemming* (merubah kumpulan kata menjadi kata dasarnya), maka kumpulan tweet menjadi kumpulan kata-kata yang dibagi menjadi kumpulan kata untuk data training pada tabel 3 dan kumpulan data testing pada tabel 4.

Tabel 3. Kumpulan Data Training

No.	Kategori	Jumlah kata
1	Positif	266
2	Negatif	562
3	Netral	330
Total		1158

Tabel 4. Kumpulan Data Testing

No.	Kategori	Jumlah kata
1	Positif	127
2	Negatif	273
3	Netral	178
Total		578

3.5 Tingkat Akurasi Pada Klasifikasi Naive Bayes

3.5.1 Tingkat Akurasi Data Training

Berdasarkan hasil analisis sentimen data training dengan WEKA menggunakan klasifikasi Naive Bayes maka dihasilkan tingkat akurasi (*accuracy*) atau tingkat kedekatan antara nilai prediksi dan nilai aktual sebesar 81,0881% sebagaimana terdapat pada gambar 3.

=== Summary ===

```
Correctly Classified Instances      939      81.0881 %
Incorrectly Classified Instances    219      18.9119 %
```

Gambar 2. Hasil Sentimen Analisis dengan Weka pada data training

Berdasarkan perhitungan tingkat akurasi pada data training maka akan dihasilkan *Confusion matrix* yaitu suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining atau Sistem

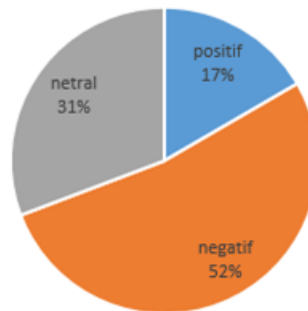


Pendukung Keputusan. Pada tabel 5 merupakan tampilan *confusion matrix* data training yang bernilai data prediksi dan data aktual.

Tabel 5. *Confusion Matrix* Data Training

		PREDIKSI			
		Positif	Negatif	Netral	
AKTUAL	Positif	266	156	59	51
	Negatif	562	9	494	59
	Netral	330	7	34	289

Dari tabel 5 dapat diketahui bahwa klasifikasi Naive Bayes berhasil melakukan klasifikasi secara benar terhadap sentimen positif sebesar 156 (17%), sentimen negatif sebanyak 494 (52%) dan sentimen netral sebanyak 289 (31%) yang digambarkan dalam bentuk grafik pada gambar 4.



Gambar 3. Hasil Analisis Sentimen pada Data Training dengan Klasifikasi Naive Bayes

3.5.2 Tingkat akurasi Data Testing

Berdasarkan kumpulan data testing dihasilkan bahwa tingkat akurasi (*accuracy*) klasifikasi Naive Bayes bernilai 76,2976% sebagaimana terdapat pada gambar 5.

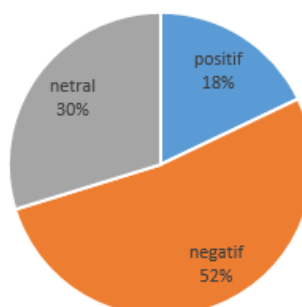
Correctly Classified Instances	441	76.2976 %
Incorrectly Classified Instances	137	23.7024 %

Gambar 4. Hasil Sentimen Analisis dengan Weka pada data testing

Perhitungan tingkat akurasi pada data testing akan menghasilkan *Confusion matrix* yang terdapat pada tabel 6 dimana klasifikasi Naive Bayes berhasil melakukan klasifikasi secara benar untuk sentimen positif sebanyak 92 (18%), sentimen negatif sebanyak 269 (52%) dan sentimen netral sebanyak 153 (30%) dan grafik ditampilkan pada gambar 6.

Tabel 6. *Confusion Matrix* Data Testing

		PREDIKSI			
		Positif	Negatif	Netral	
AKTUAL	Positif	127	92	19	16
	Negatif	273	2	269	2
	Netral	178	2	23	153



Gambar 5. Hasil Analisis Sentimen pada Data Testing dengan Klasifikasi Naive Bayes



3.5.3 Evaluasi Klasifikasi Model

Untuk mengetahui klasifikasi model sudah benar maka berikut ini adalah tampilan pembuktian dari WEKA untuk data training yang ditampilkan pada gambar 7 dan tampilan pembuktian data testing pada gambar 8.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,876	0,133	0,724	0,876	0,793	0,706	0,933	0,884	Netral
	0,879	0,156	0,842	0,879	0,860	0,723	0,954	0,953	Negatif
	0,586	0,018	0,907	0,586	0,712	0,672	0,945	0,769	Positif
Weighted Avg.	0,811	0,118	0,823	0,811	0,807	0,706	0,946	0,891	

Gambar 6. Tampilan Laporan Klasifikasi Naive Bayes dengan Data Training

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,985	0,377	0,701	0,985	0,819	0,643	0,978	0,978	Negatif
	0,150	0,009	0,826	0,150	0,253	0,298	0,972	0,862	Positif
	0,860	0,045	0,895	0,860	0,877	0,824	0,975	0,946	Netral
Weighted Avg.	0,763	0,194	0,788	0,763	0,712	0,623	0,976	0,943	

Gambar 7. Tampilan Laporan Klasifikasi Naive Bayes Dengan Data Testing

Berdasarkan gambar 7 dan 8 maka akan dibuatkan tabel kesimpulan yang terdapat pada tabel 7.

Tabel 7. Tingkat Evaluasi Model dengan Klasifikasi Naive Bayes

No	Data	Positif	Negatif	Netral
Data Training				
1	Precision	0,907	0,842	0,724
2	Recall	0,586	0,879	0,876
Data Testing				
1	Precision	0,826	0,701	0,895
2	Recall	0,150	0,985	0,860
Rata-rata				
1	Precision	0,8665	0,7715	0,8095
2	Recall	0,368	0,932	0,868

Dari Tabel 7 diketahui bahwa data training mendapatkan nilai presisi (*precision*) [15] yaitu tingkat ketepatan antara kata yang sudah diberikan label dengan klasifikasi yang diberikan oleh Naive bayes dengan nilai sentimen positif adalah 0,907 atau 90,7%, sentimen Negatif adalah 0,842 atau 84,2%, sedangkan sentimen Netral adalah 0,724 atau 72,4%. Sedangkan pada data testing memiliki nilai presisi (*precision*) untuk sentimen positif adalah 0,826 atau 82,6%, sentimen negatif 0,701 atau 70,1% dan sentimen netral 0,895 atau 89,5%.

Pada tabel 7 juga terdapat nilai *recall* [15] yaitu tingkat keberhasilan klasifikasi Naive Bayes dalam menemukan kembali sebuah informasi, informasi pada penelitian ini berarti kata yang sudah diberikan label sentimen, maka berdasarkan tabel 7 nilai *recall* untuk data training untuk sentimen positif bernilai 0,586 atau 58,6%, sentimen negatif bernilai 0,879 atau 87,9% dan sentimen netral bernilai 0,876 atau 87,6%. sedangkan nilai recall untuk data testing adalah sentimen positif bernilai 0,150 atau 15,0%, sentimen negatif bernilai 0,985 atau 98,5% dan sentimen netral bernilai 0,860 atau 86,0%.

3.5.4 Ketepatan Klasifikasi Naive Bayes terhadap Data Tweet

Berdasarkan gambar 7 dan 8 masih berkaitan dengan evaluasi model dengan klasifikasi naive bayes maka akan didapatkan nilai ROC AREA [15] yaitu kemampuan tes untuk mengklasifikasikan dengan benar yang dirangkum pada tabel 8.

Tabel 8. Nilai ROC AREA pada Klasifikasi Naive Bayes

No	Data	Positif	Negatif	Netral
1	Training	0,945	0,954	0,933
2	Testing	0,972	0,978	0,975

Nilai ROC Area antara 0,90-1 bernilai Excellent (A) [15] sehingga dari Tabel 9 dapat disimpulkan bahwa klasifikasi Naive Bayes terbukti dapat mengklasifikasi dengan tepat sentimen analisis untuk melihat persepsi masyarakat di Indonesia terhadap pemilu 2019 menggunakan data media sosial Twitter sebanyak 221.



4. KESIMPULAN

Berdasarkan analisis dan pembahasan maka kesimpulan penelitian ini adalah : (1) Data tweet dikumpulkan dari tanggal 16 April 2018 sampai 16 April 2019 sehingga didapatkan data sebanyak 221 data tweet yang kemudian diberi label oleh 2 orang tenaga ahli di bidang komunikasi untuk menentukan sentimen positif, negatif dan netral yang kemudian dibagi menjadi data training 70% atau sebanyak 154 data dan data testing 30% atau sebanyak 67 data setelah itu dilakukan proses preprocessing dan text transformation dari proses tersebut terbentuklah data set pola persepsi yang dapat digunakan untuk analisis sentimen menggunakan modeling klasifikasi dengan Naive Bayes. (2) Berdasarkan data set yang telah terbentuk pola persepsi maka dilakukan proses modeling dengan klasifikasi Naive Bayes dan menemukan bahwa data set tweet pada penelitian memiliki pola persepsi negatif atau sentimen negatif sebesar 52% jauh lebih besar dari persepsi positif yaitu 18% dan Persepsi positif juga lebih kecil dari persepsi netral dimana persepsi netral memiliki nilai 31%. Tingkat akurasi klasifikasi Naive Bayes terhadap dataset training bernilai 81% dan 76% pada dataset testing, sedangkan rata-rata nilai precision untuk sentimen positif 86,65%, sentimen negatif 77,15%, dan sentimen netral bernilai 80,95% sedangkan rata-rata nilai recall pada sentimen positif 36,8%, sentimen negatif 93,2% dan sentimen netral 86,8%.

REFERENCES

- [1] S. Sorik, "Penataan Demokrasi dan Pemilu di Indonesia Pasca Reformasi," *J. Penelit. Polit.*, vol. 16, no. 1, pp. 101–107, 2019.
- [2] H. Abdulsalam, "Pemilu 2019 Dibayangi Ancaman Golput demi Liburan Baca," *tirto.id*, 2019. [Online]. Available: <https://tirto.id/pemilu-2019-dibayangi-ancaman-golput-demi-liburan-djex>. [Accessed: 04-Apr-2019].
- [3] E. Sulisty, "Partisipasi Pemilih dalam Pemilu," *Koran Sindo*, 05-Mar-2019.
- [4] A. Akbar et al., *Serial Evaluasi Penyelenggaraan Pemilu Serentak 2019 Perihal Partisipasi Masyarakat*. Jakarta: Badan Pengawas Pemilihan Umum (BAWASLU) RI, 2019.
- [5] B. Clinten, "Pengguna Aktif Harian Twitter Indonesia Di Klaim Terbanyak," *Kompas.com*, 2019. [Online]. Available: <https://tekno.kompas.com/read/2019/10/30/16062477/pengguna-aktif-harian-twitter-indonesia-diklaim-terbanyak>. [Accessed: 30-Dec-2019].
- [6] S. J. Pan, X. Ni, J. T. Sun, Q. Yang, and Z. Chen, "Cross-Domain Sentiment Classification via Spectral Feature Alignment," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 751–760.
- [7] H. Simorangkir and K. M. Lhaksmana, "Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naive Bayes Classifier," *e-proceeding of Englineering*, vol. 5, no. 3, pp. 8131–8140, 2018.
- [8] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [9] C. Prianto, N. H. Harani, and I. Firmansyah, "Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter," *J. Media Inform. Budidarma*, vol. 3, no. 4, pp. 405–413, 2019.
- [10] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019.
- [11] P. Y. Saputra, "Implementasi Teknik Crawling Untuk Pengumpulan Data Dari Media Sosial Twitter," *J. Din. Dotcom*, vol. 8, no. 2, pp. 160–168, 2017.
- [12] T. T. Sang Nguyen, "Model-Based Book Recommender Systems using Naive Bayes enhanced with Optimal Feature Selection," in *ACM International Conference Proceeding Series*, 2019, no. February, pp. 217–222.
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [14] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 4th ed. United States: Morgan Kaufmann, 2016.
- [15] I. H. Witten, E. Frank, M. A. Hall, and C. J Pal, *The WEKA Workbench online appendix*. New Zealand: The University of Waikato, 2016.