

Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen *Review Data Twitter BMKG Nasional*

Dedi Darwis^{1,*}, Nery Siskawati¹, Zaenal Abidin¹

¹ Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
^{*}darwisdedi@teknokrat.ac.id

Abstrak— Pertumbuhan *Twitter* terus meningkat setiap waktu, sehingga hal tersebut dimanfaatkan para pengguna *Twitter* untuk menyampaikan informasi berupa kritik maupun saran kepada pelayanan yang diberikan BMKG Nasional dengan lebih mudah. Metode yang digunakan dalam penelitian ini adalah klasifikasi data adalah *Naive Bayes Classifier* (NBC). Sistem yang dikembangkan dengan menggunakan data internal yang diambil dari internet/*Twitter* untuk proses penentuan kalimat termasuk opini positif, netral atau negatif. Penentuan tersebut digolongkan sebagai proses pengklasifikasian. Serta menggunakan *Application Python 3.74*. Hasil Penelitian ini masuk kedalam *fined grained sentiment analysis* yaitu analisis pada suatu kalimat komentar. Data tersebut akan diproses menggunakan *text mining*, kemudian dilanjutkan dengan mengklasifikasikan *Tweet* ke dalam tiga kelas, yaitu positif, negatif, dan netral. Klasifikasi ini menggunakan algoritma *naive bayes*. Klasifikasi dapat memberikan kemudahan bagi pengguna untuk melihat opini positif, negatif, dan netral. Hasil uji akurasi pada metode *naive bayes* untuk klasifikasi yaitu 69.97%.

Kata Kunci: Analisis Sentimen, *Naive Bayes*, *Python 3.74*, *Twitter*.

Abstract— *Twitter growth continues to increase every time, so it is used by Twitter users to convey information in the form of criticism and suggestions to the services provided by the National BMKG more easily. The method used in this study is the classification of data is the Naive Bayes Classifier (NBC). The system was developed using internal data taken from the internet / Twitter for the process of determining sentences including positive, neutral or negative opinions. The determination is classified as a classification process. And using Python 3.74 application. The results of this study are included in the fined grained sentiment analysis, which is an analysis in a comment sentence. The data will be processed using text mining, then proceed with classifying Tweets into three classes, namely positive, negative, and neutral. This classification uses the Naive Bayes algorithm. Classification can make it easy for users to see positive, negative and neutral opinions. The results of the accuracy test on the Naive Bayes method for classification are 69.97%.*

Keywords: Sentiment Analysis, *Naive Bayes*, *Python 3.74*, *Twitter*.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi internet saat ini membawa dampak perubahan pada beberapa sektor kehidupan manusia, seperti hanya pada bidang pendidikan, perdagangan, pemerintahan hingga komunikasi social. Di era sekarang ini pertumbuhan internet dan media sosial telah berkembang pesat sehingga menyediakan berbagai informasi opini dari orang lain. Sebuah *survey* yang dilakukan oleh Asosiasi Pengusaha Jasa Internet (APJII) pada tahun 2014 menunjukkan bahwa jumlah pengguna internet di Indonesia mengalami kenaikan dalam beberapa tahun terakhir. Pada tahun 2005, tercatat sebanyak 16 juta orang pengguna internet di Indonesia dan angka ini bertambah menjadi 20 juta orang pengguna internet pada tahun 2006. Penelitian ini juga memberikan hasil bahwa saat ini penggunaan internet paling banyak adalah akses ke situs media sosial (87,4% dari total responden), salah satu media social pada zaman sekarang yaitu *Twitter* [1].

Media sosial *Twitter* adalah salah satu sebagai media penghubung yang diminati oleh seluruh masyarakat di dunia [2]. Hal ini dapat dilihat dari jumlah peningkatan pengguna *Twitter* yang tercatat di seluruh dunia salah satunya Indonesia. Pada saat ini *Twitter* memiliki jumlah pengguna aktif sebesar 313 juta per bulan pada tahun 2016 [3]. Pengguna *Twitter* akan memberikan kabar terbaru atau komentar tentang hal yang sedang menjadi topik utama di dunia. Hal yang sedang menjadi topik utama dan banyak dan sering dikomentari oleh pengguna akan menimbulkan suatu masalah atau trending topik di media sosial terutama *Twitter*.

Pertumbuhan *Twitter* terus meningkat setiap waktu [4], sehingga hal tersebut dimanfaatkan para pengguna *Twitter* untuk menyampaikan informasi berupa komentar kritik maupun saran kepada pelayanan yang diberikan BMKG Nasional dapat dilakukan dengan lebih mudah. Semakin banyak pendapat atau keluhan dari beberapa masyarakat dalam membentuk opini masyarakat, dan dapat dijadikan masukan terhadap suatu penilaian pelayanan yang diberikan BMKG Nasional. Pada pelayanan yang diberikan oleh salah satu petugas BMKG Nasional yang kurang baik maka akan menimbulkan banyak opini masyarakat tentang pelayanan BMKG Nasional. Analisis sentimen dapat membantu untuk memperoleh gambaran umum persepsi masyarakat dengan

cara mengelompokkan jenis opini menjadi kategori positif, negatif, atau netral terhadap pelayanan pada BMKG Nasional.

Algoritma *Naive Bayes* adalah merupakan salah satu algoritma klasifikasi yang banyak digunakan pada Data Mining ataupun Text Mining [5]. Algoritma *Naive Bayes* berdasarkan teorema *Bayes* bahwa semua kegiatan memberikan sebuah kontribusi yang sama penting atau saling bebas pada pemilihan kelas tertentu. Salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam Text Mining adalah metode *Naive Bayes* yang sering disebut dengan *Naive Bayes Classifier* [6], [7].

Penelitian yang dilakukan adalah tentang sentimen analisis untuk mengklasifikasikan *Tweet Review* BMKG Nasional. Penelitian ini termasuk kedalam sebuah penelitian yang bernama *fined grained sentiment analysis* yaitu analisis pada suatu kalimat komentar. Data dari *Twitter* tersebut akan dapat diproses menggunakan *text mining*, kemudian dilanjutkan dengan mengklasifikasikan *Tweet* ke dalam tiga kelas, yaitu positif, negatif, dan netral. Klasifikasi dapat memberikan kemudahan bagi semua pengguna untuk melihat kritik dan saran [8], [9]. Tujuan penelitian ini adalah untuk melakukan pencarian komentar negatif, positif, dan netral dengan review data *Twitter* BMKG Nasional menggunakan algoritma *naive bayes*, lalu melakukan perhitungan data negatif, positif, dan netral menggunakan algoritma *naive bayes*, serta dapat menghasilkan klasifikasi berdasarkan nilai yang didapat dengan algoritma *naive bayes*. Penelitian yang dilakukan difokuskan kepada kritik dan saran dalam pelayanan BMKG Nasional yang ada di akun resmi *Twitter* dengan menggunakan data pada tahun 2018 dan 2019. Proses pengambilan data dan klasifikasi sentimen analisis menggunakan bahasa pemrograman Python 3.74.

1.2 Landasan Teori

a. Analisis Sentiman

Analisis Sentimen adalah salah satu cabang sebuah penelitian text mining, yang berkaitan dengan bidang yang lebih luas seperti pengolahan data kegiatan tertentu [3].

b. Text Mining

Text Mining merupakan tahapan proses dari analisis dalam data yang berupa teks dimana sumber data didapatkan dari suatu dokumen seperti kalimat data kata, Konsep text mining biasanya digunakan dalam klasifikasi dokumen tekstual dimana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen tersebut [9].

c. Naive Bayes

Salah satu tugas Data Mining adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu kelas atau beberapa kelas yang sudah didefinisikan sebelumnya. Salah satu metoda dalam klasifikasi data adalah *Naive Bayes* yaitu salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, cara kerja *Naive Bayes* yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [7]. Dasar dari *Naive Bayes* yang dipakai dalam pemrograman adalah rumus Bayes seperti pada persamaan (1)

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

Keterangan :

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B.

Pada pengaplikasiannya nanti rumus pada persamaan (1) berubah menjadi persamaan (2).

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

Naive Bayes atau bisa disebut sebagai *Multinomial Naive Bayes* merupakan model penyederhanaan dari Metoda Bayes yang cocok dalam pengklasifikasian teks atau dokumen.

Pada persamaan (3) merupakan persamaan model penyederhanaan dari Metoda Bayes.

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

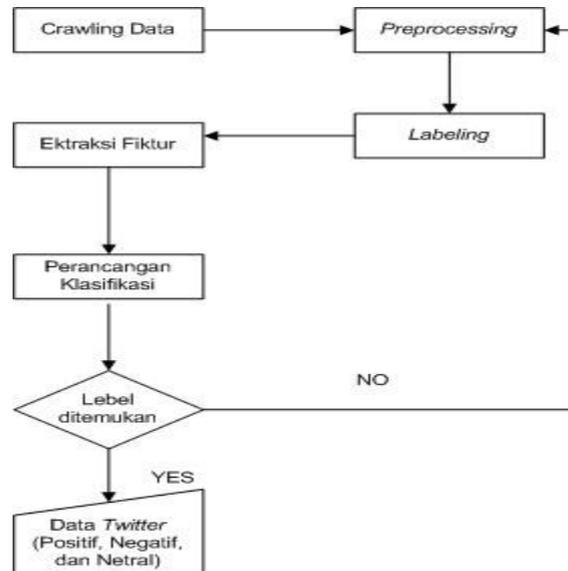
Berdasarkan persamaan (3), maka persamaan (1) dapat ditulis seperti yang terdapat pada persamaan (4)

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

2. METODE PENELITIAN

2.1 Kerangka Penelitian

Kerangka penelitian dibuat untuk membuat pola pikir penelitian menjadi lebih realistik. Pada penelitian yang dilakukan, gambaran tentang kerangka penelitian yang dibuat disajikan pada Gambar 1.



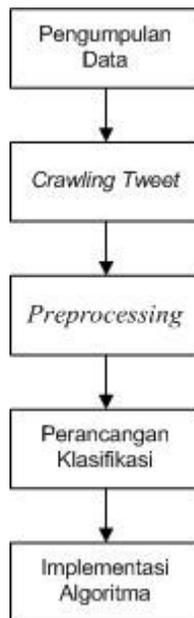
Gambar 1. Kerangka Penelitian

Penjelasan pada Gambar 1.

- Crawling Data*, yaitu data *Twitter* berisi opini masyarakat yang mengandung unsur komentar positif, netral, dan negatif pada BMKG dengan data tahun 2018-2019 melalui aplikasi *Twitter*. Pada proses ini dimulai dari melakukan kunjungan *Twitter* BMKG kemudian menuju ke folder *clone* dan mulai melakukan penyaringan data dengan *Python*.
- Pelabelan sentimen *Tweet* menjadi netral, positif atau negatif
- Preprocessing*, dilakukan untuk menghasilkan data bersih. *Preprocessing* meliputi: *casefolding*, *filtering*, *tokenisasi*, *slang replacement* dan *stopwordremoval*.
- Ekstraksi Fitur, meliputi: *Unigram*, *Negation*, *TF* dan *TF-IDF*.
- Perancangan klasifikasi, dilakukan untuk mengelompokkan *Tweet* berdasarkan kelas yang ditentukan.
- Label ditemukan, digunakan untuk menemukan klasifikasi data baru menggunakan *Naïve Bayes*.

2.2 Tahapan Penelitian

Tahapan penelitian dalam pengambilan data sentimen analisis berupa *Tweet* negatif, positif, dan netral pada *Twitter* BMKG, dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Penelitian

2.3 Cara Kerja Sentimen Analisis

Proses klasifikasi menggunakan data pada tahun 2018-2019 pada *Twitter* BMKG. *Tweet* dibagi ke dalam tiga kelas kata, yaitu positif, netral dan negatif. Setiap dokumen yang diklasifikasi akan di kategorikan ke dalam kelas yang paling dominan, berikut ini adalah tahapan yang dilakukan :

a. *Crawler Twitter* Akun BMKG

Crawler merupakan sebuah alat untuk mengindeks dan mengunduh konten dari internet, lalu disimpan ke dalam database mesin pencari [10]. Berikut ini adalah hasil crawl *Twitter* akun BMKG salah satu dari komentar yang ada pada akun BMKG, sebagai contoh yaitu :

ari10prasetyo @ BMKG @Awas aja kalau alat kayak gini masih ada :)

b. *Case Folding*

Case folding merupakan proses untuk mengubah kalimat atau menjadi huruf kecil [11]. Berikut ini adalah contoh hasil *Case Folding* akun BMKG salah satu dari komentar yang ada pada akun BMKG yaitu :

ari10prasetyo @ bmkg @awas aja kalau alat kayak gini masih ada :)

c. *Cleansing*

Data cleansing adalah kegiatan menganalisa kualitas data dengan cara memodifikasi, mengubah, atau menghapus data-data yang dianggap tidak perlu, tidak lengkap, data tidak akurat, atau memiliki format data atau file yang salah dalam basis data guna menghasilkan data berkualitas tinggi. *Data cleansing* juga biasa disebut *data cleaning* atau *data scrubbing* [12]. Berikut ini adalah contoh *Cleansing* akun dengan mencantumkan komentar yang dilakukan tanpa adanya nama pengomentor.

awas aja kalau alat kayak gini masih ada :)

d. *Stopword Removal*

Stopword removal disebut juga *filtering*, adalah tahap pemilihan kata-kata penting dari hasil token, yaitu kata-kata apa saja yang akan digunakan untuk mewakili dokumen [13]. Berikut ini adalah *stopword removal* yang dilakukan yaitu proses penyatuan dan pergantian emot yang dilakukan oleh pengomentor, dapat dilihat dibawah ini :

: Alat kayak gini masih ada, tapi emot senang

e. *Tokenisasi*

Tokenisasi adalah proses memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau *termmed word* yang berdiri sendiri [14]. Berikut ini adalah contoh proses *tokenisasi* yaitu penyaringan kata yang dilakukan oleh pengomentor.

Alat, kayak, gini, masih ada

2.4 Metode Pengumpulan Data

Pada penelitian ini digunakan data sekunder yang diambil dari akun *Twitter* resmi BMKG Nasional. Data yang diambil akan dijadikan sebagai data training untuk membuat kamus opini. Pembuatan kamus data opini tersebut dengan mengumpulkan data opini melalui media data online. Kamus data opini yang dikumpulkan diklasifikasi menjadi kamus opini negatif dan positif. Untuk melakukan klarifikasi peneliti menggunakan bantuan secara online maupun offline untuk melakukan klarifikasi agar data kamus sesuai dan meminimalkan subjektivitas.

Aplikasi yang dikembangkan menggunakan data internal yang diambil dari internet/*Twitter* untuk proses penentuan kalimat termasuk opini positif, netral atau negatif. Penentuan tersebut digolongkan sebagai proses pengklasifikasian. Pengklasifikasian pada suatu dokumen termasuk kelas positif, kelas netral atau kelas negatif. Didalam pengumpulan data ini juga menggunakan study literature dengan data sekunder yang didapat dari internet/*Twitter*.

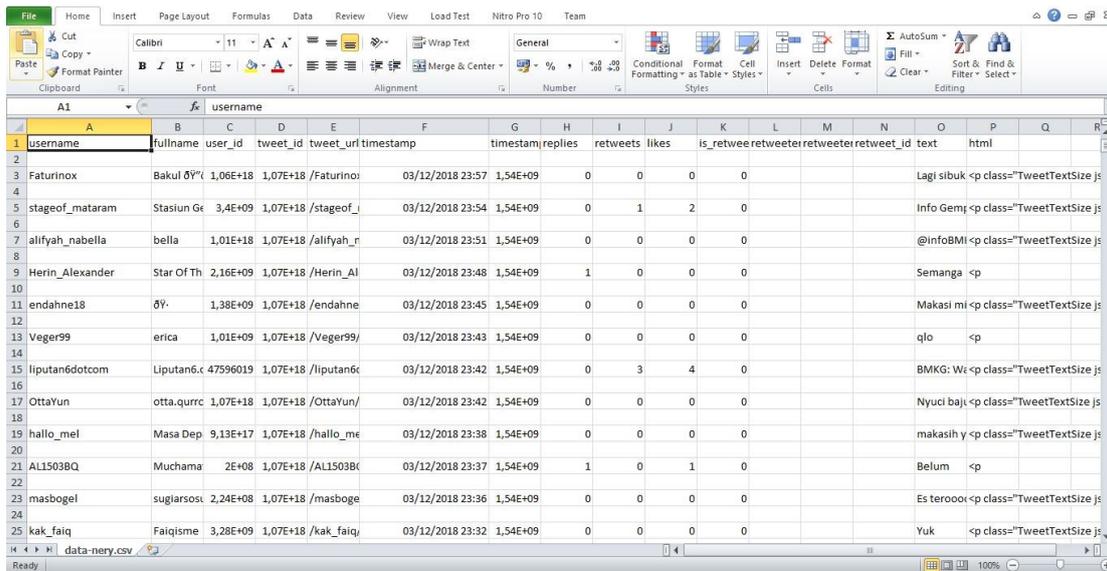
2.5 Proses *Twitter* Crawling

Proses crawling pada *Twitter* dilakukan dengan memanfaatkan fasilitas yang ada pada bahasa pemrograman Python 3.74 yang telah disediakan. Dengan terlebih dahulu melakukan pengambilan data aplikasi pada *Twitter* Developer. Analisis sentiment pada penelitian ini dibatasi pada BMKG, dan menghasilkan penyimpanan data dalam format CSV. Gambar 2 merupakan potongan source code untuk mengambil data pada *Twitter*.

```
- Python -version
- Python -m pip install twitterscraper
- twitterscraper "BMKG" --output data-nery.csv -
  -limit 1000 --begindate 2018-9-4 --enddate
  2019-9-4 --csv
```

Gambar 3. Potongan Source Code Crawling Data *Twitter*

Dari data hasil crawling selanjutnya disimpan pada file dengan format CSV. Gambar 4 menunjukkan contoh data yang akan disimpan untuk proses pembentukan model.



Gambar 4. Hasil Crawling Data *Twitter* dalam Format CSV

2.6 Skenario Pengujian

Pada skenario pengujian ini jumlah data yang digunakan adalah 1179 data *Tweet*. Data dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan perbandingan 70:825 dimana data yang digunakan tersebut sudah diberi label. Oleh algoritma Naive Bayes Classifier(NBC), data *training* digunakan untuk membentuk tabel probabilitas, dan data *testing* digunakan untuk menguji tabel probabilitas yang telah terbentuk atau dapat diasumsikan bahwa data *training* adalah data yang digunakan sebagai acuan untuk membangun model klasifikasi, sedangkan data *testing* adalah data yang digunakan untuk menguji performa dari model klasifikasi tersebut.

Pada skenario pengujian ini data *training* sudah dipersiapkan sebelumnya dan disimpan ke dalam dataset. Hasil data *training* (pembelajaran) yang didapat akan diuji dengan data baru yang masih belum diketahui *classtype* kategorinya. Pada penelitian ini terdapat proses *preprocessing*. Adapun tahapan ini adalah *Normalisasi fitur*, *tokenizing*, *stop-words*, dan *stemming*. Proses ini bertujuan untuk mendapatkan data *training* yang sesuai harapan si analis. Disini penulis melakukan variasi pada tahap *preprocessing* dengan menghilangkan salah satu tahapan *preprocessing* yaitu tahap Normalisasi fitur. Pada tahapan ini penulis ingin mengetahui tingkat akurasi ketika salah satu tahapan dihilangkan apakah akan berpengaruh pada hasil akurasi atau tidak. Tabel 1 merupakan pembagian data training dan data testing dalam proses pengujian.

Tabel 1. Pembagian Data *Training* dan Data *Testing*

No	Class	Data Training	Data Testing
1	Positif	931	354
2	Negatif	70	354
3	Netral	178	354
	Total	1179	1179

3. HASIL DAN PEMBAHASAN

3.1 Menghitung Probabilitas Kata

Menghitung probabilitas kata dilakukan untuk bertujuan mendapatkan term dengan nilai yang lebih penting dan dianggap relevan untuk dijadikan kata kunci. Proses pembobotan menggunakan algoritma Naive Bayes Classifier dalam proses perhitungan persamaan dapat dilihat pada Tabel 2.

Tabel 2. Contoh Kemunculan Term Frekuensi

Kata	Frekuensi kemunculan kata (W_k)		
	354 <i>Tweet</i> Positif	354 <i>Tweet</i> Negatif	354 <i>Tweet</i> Netral
Pagi	189	7	-
Cuaca	7	3	5
Baik	13	1	6
Oke	5	3	-
Hoak	1	1	-
Jangan	6	1	1
Salah	3	-	-
Bantu	2	-	1
Mungkin	2	5	-
Tumben	-	2	-

Selanjutnya mencari probabilitas kata pagi, cuaca, baik, oke, hoak, jangan, salah, bantu, mungkin, tumben:
Diketahui :

n_{Tweet} Positif : 354
 n_{Tweet} Negatif : 354
 n_{Tweet} Netral : 354

a. Kata Pagi

$$P(\text{Pagi}|\text{Positif}) = \frac{189+1}{354+825} = 0,16$$

$$P(\text{Pagi}|\text{Negatif}) = \frac{7+1}{354+825} = 0,006$$

$$P(\text{Pagi}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008$$

b. Kata Cuaca

$$P(\text{Cuaca}|\text{Positif}) = \frac{7+1}{354+825} = 0,006$$

$$P(\text{Cuaca}|\text{Negatif}) = \frac{3+1}{354+825} = 0,003$$

$$P(\text{Cuaca}|\text{Netral}) = \frac{5+1}{354+825} = 0,005$$

c. Kata Baik

$$P(\text{baik}|\text{Positif}) = \frac{13+1}{354+825} = 0,011$$

$$P(\text{baik}|\text{Negatif}) = \frac{1+1}{354+825} = 0,0016$$

$$P(\text{baik}|\text{Netral}) = \frac{6+1}{354+825} = 0,005$$

d. Kata oke

$$P(\text{Oke}|\text{Positif}) = \frac{5+1}{354+825} = 0,005$$

$$P(\text{Oke}|\text{Negatif}) = \frac{3+1}{354+825} = 0,003$$

$$P(\text{Oke}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008$$

e. Kata Hoak

$$P(\text{Hoak}|\text{Positif}) = \frac{1+1}{354+825} = 0,0016$$

$$P(\text{Hoak}|\text{Negatif}) = \frac{1+1}{354+825} = 0,0016$$

$$P(\text{Hoak}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008$$

f. Kata Jangan

$$P(\text{jangan}|\text{Positif}) = \frac{6+1}{354+825} = 0,0059$$

$$P(\text{jangan}|\text{Negatif}) = \frac{6+1}{354+825} = 0,0059$$

$$P(\text{jangan}|\text{Netral}) = \frac{1+1}{354+825} = 0,0016$$

g. Kata Salah

$$P(\text{Salah}|\text{Positif}) = \frac{3+1}{354+825} = 0,003$$

$$P(\text{Salah}|\text{Negatif}) = \frac{0+1}{354+825} = 0,0008$$

$$P(\text{Salah}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008$$

h. Kata Bantu

$$P(\text{Bantu}|\text{Positif}) = \frac{2+1}{354+825} = 0,0025$$

$$P(\text{Bantu}|\text{Negatif}) = \frac{0+1}{354+825} = 0,0008$$

$$\begin{aligned}
 & \frac{354+825}{354+825} \\
 \text{i. Kata Mungkin} & \quad P(\text{Bantu}|\text{Netral}) = \frac{1+1}{354+825} = 0,0016 \\
 & \quad P(\text{Mungkin}|\text{Positif}) = \frac{2+1}{354+825} = 0,0025 \\
 & \quad P(\text{Mungkin}|\text{Negatif}) = \frac{5+1}{354+825} = 0,005 \\
 & \quad P(\text{Mungkin}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008 \\
 \text{j. Kata Tumben} & \quad P(\text{Tumben}|\text{Positif}) = \frac{0+1}{354+825} = 0,0008 \\
 & \quad P(\text{Tumben}|\text{Negatif}) = \frac{2+1}{354+825} = 0,0025 \\
 & \quad P(\text{Tumben}|\text{Netral}) = \frac{0+1}{354+825} = 0,0008
 \end{aligned}$$

Setelah mendapatkan hasil dari probabilitas kata, kemudian akan menghitung probabilitas dari dokumen (*Tweet*) sampel atau contoh. Diasumsikan $P(V_j)$ (probabilitas kategori dokumen) sama dengan $\frac{docs_j}{|contoh|}$ (jumlah dokumen setiap kategori) dibagi dengan $|contoh|$ (jumlah dokumen yang digunakan sebagai data training dari seluruh kategori). Diperoleh persamaan (5).

$$P(v_j) = \frac{|docs_j|}{|contoh|} \quad (5)$$

Perhitungan probabilitas *Tweet* dimana terdapat 1179 *Tweet* yang terbagi 354 *Tweet* kelas positif, 354 *Tweet* kelas negatif dan 354 *Tweet* kelas netral :

$$\begin{aligned}
 P(\text{Positif}) &= \frac{354}{1179} = 0,3 \\
 P(\text{Negatif}) &= \frac{354}{1179} = 0,3 \\
 P(\text{Netral}) &= \frac{354}{1179} = 0,3
 \end{aligned}$$

Data kata dari *Tweet* nilai probabilitas dapat dilihat pada Tabel 3.

Tabel 3. Daftar Uji Kata

No	Kata	Prob positif V1	Prob Negatif V2	Prob Netral V3
1	Pagi	0,16	0.006	0,0008
2	Cuaca	0,006	0,003	0,005
3	Baik	0,0011	0,0016	0,005
4	Oke	0,005	0,003	0,0008
5	Hoak	0,0016	0,0016	0,0008
6	Jangan	0,0059	0,0059	0,0016
7	Salah	0,003	0,0008	0,0008
8	Bantu	0,0025	0,0008	0,0016
9	Mungkin	0,0025	0,005	0,0008
10	Tumben	0,0008	0,0025	0,0008

Untuk mendapatkan nilai probabilitas nilai tertinggi langkah pertama yakni menghitung $(P(W_{kl}|V_j)P(V_j))$. $P(V_j)$ didapat probabilitas dokumen:

Kata (Pagi Positif)	: 0,16 *0,3	= 0,048
(Pagi Negatif)	: 0,006 *0,3	= 0,0018
(Pagi Netral)	: 0,0008 *0,3	= 0,00024
(Cuaca Positif)	: 0,006 *0,3	= 0,0018
(Cuaca Negatif)	: 0,003 *0,3	= 0,0009
(Cuaca Netral)	: 0,005 *0,3	= 0,0015
(Baik Positif)	: 0,011*0,3	= 0,0033
(Baik Negatif)	: 0,0016 *0,3	= 0,00048
(Baik Netral)	: 0,0045*0,3	= 0,00135
(Ok Positif)	: 0,005*0,3	= 0,0015
(Ok Negatif)	: 0,003 *0,3	= 0,0009
(Ok Netral)	: 0,0008 *0,3	= 0,00024
(Hoak Positif)	: 0,0016 *0,3	= 0,00048
(Hoak Negatif)	: 0,0016 *0,3	= 0,00048
(Hoak Netral)	: 0,0008 *0,3	= 0,00024
(Jangan Positif)	: 0,0059 *0,3	= 0,0017
(Jangan Negatif)	: 0,0059*0,3	= 0,0017
(Jangan Netral)	: 0,0016 *0,3	= 0,0048
(Salah Positif)	: 0,003 *0,3	= 0,0009
(Salah Negatif)	: 0,0008 *0,3	= 0,00024
(Salah Netral)	: 0,0008 *0,3	= 0,00024
(Bantu Positif)	: 0,0025 *0,3	= 0,00075
(Bantu Negatif)	: 0,0008 *0,3	= 0,00024
(Bantu Netral)	: 0,0016*0,3	= 0,00048
(Mungkin Positif)	: 0,0025*0,3	= 0,00075
(Mungkin Negatif)	: 0,005 *0,3	= 0,0015
(Mungkin Netral)	: 0,0008 *0,3	= 0,00024
(Tumben Positif)	: 0,0008 *0,3	= 0,00024
(Tumben Negatif)	: 0,0025 *0,3	= 0,00075
(Tumben Netral)	: 0,0008 *0,3	= 0,00024

Setelah itu menghitung nilai probabilitas tertinggi dari masing-masing kategori.

Probabilitas Positif Tertinggi = $0,048 * 0,0018 * 0,0033 * 0,0015 * 0,00048 * 0,0017 * 0,0009 * 0,00075 * 0,00075 * 0,00024 = 4.24019059e-29$

Probabilitas Negatif Tertinggi = $0,0018 * 0,0009 * 0,00048 * 0,000 * 0,00048 * 0,00017 * 0,00024 * 0,00024 * 0,0015 * 0,00075 = 3.70052997e-33$

Probabilitas Netral Tertinggi = $0,00024 * 0,0015 * 0,00135 * 0,0002 * 0,00024 * 0,00048 * 0,00024 * 0,00048 * 0,00024 * 0,00024 = 8.91610045e-35$

Tabel 4. Nilai Probabilitas

Probabilitas Positif Tertinggi	Probabilitas Negatif Tertinggi	Probabilitas Netral Tertinggi
4.24019059e-29	3.70052997e-33	8.91610045e-35

Dari perhitungan algoritma *naive bayes* yang dilakukan, kemudian dicari perbandingan nilai probabilitas tertinggi dari masing-masing *Tweet* sehingga *Tweet* yang sudah di *testing* dapat di kategorikan kedalam dokumen *Tweet* opini yang sesuai dengan isi teksnya. Pada tabel 4 dapat dilihat hasil dari keseluruhan proses perhitungan probabilitas tertinggi dengan *naive bayes*. Pada tabel tersebut probabilitas positif memiliki nilai tertinggi. Sehingga dapat dipastikan *Tweet* yang dipilih merupakan dokumen *Tweet* positif.

3.2 Pembahasan

3.2.1 Hasil Pembagian Data

Berikut adalah rekapan hasil pembagian data positif, netral, dan negatif dapat dilihat pada Tabel 5, 6, 7 dan 8.

Tabel 5. Kasus Data *Training* Positif

<i>Tweet</i>	Kalimat	Label
<i>Tweet1</i>	Dengarkan update informasi	Positif
<i>Tweet2</i>	baik, tunggu. terimakasih banyak	Positif
<i>Tweet3</i>	silakan kami sudah respon pesan terimakasih	Positif
<i>Tweet4</i>	Peringatan dini cuaca wilayah Kalimantan Tengah	Positif
<i>Tweet5</i>	Selamat Pagi, mohon informasi yang akurat, terima kasih.	Positif
<i>Tweet6</i>	Cuaca di Jakarta hari ini (23/6) diperkirakan cerah dan berawan sepanjang hari.	Positif
<i>Tweet7</i>	Bahkan BMKG gak bikin peringatan dini buat Kepulauan Bangka Belitung	Positif
<i>Tweet8</i>	Utara Jawa dong min	Positif
<i>Tweet9</i>	Prediksi BMKG Jogja bisa mencapai suhu 8Â°c	Positif
<i>Tweet10</i>	Aku selalu bilang sama oppung, ada terpasang info bmkpung di hpku jd tau nnti kalo ada peringatan, tenang aja oppung. Bbrp bulan lalu, tiba tiba laut surut jauh. Seingatku kami ngungsi akhirnya sampe jam 5	Positif

Tabel 6. Kasus Data *Training* Negatif

<i>Tweet</i>	Kalimat	Label
<i>Tweet1</i>	Awas aja, kalo alat ini masih ada	Negatif
<i>Tweet2</i>	Sini aju naik, sini aku naikin gajinya	Negatif
<i>Tweet3</i>	Setiap nebak pasti ditolak	Negatif
<i>Tweet4</i>	#Gempa Mag:3.2, 16-Nov-18 02:00:36 WIB, Lok:2.99 LS, 119.50 BT (Pusat gempa berada di darat 21 km Tenggara Mamasa), Kedlmn: 10 Km Dirasakan (MMI) III Mamasa #BMKG...	Negatif
<i>Tweet5</i>	Dirasakan di Lombok Barat juga	Negatif
<i>Tweet6</i>	RT BpbdTrenggalek "Prakiraan cuaca Kabupaten Trenggalek	Negatif
<i>Tweet7</i>	Yaudah sih, lagian ini juga BNPB bukan BMKG yg ngasih Info.	Negatif
<i>Tweet8</i>	Sumpah gue lagi tidur semalem kaya ada yg nyenggol jadi parno eh tau tau gempa	Negatif
<i>Tweet9</i>	Emang due kepala kok sok2an sambat ngelu ?	Negatif
<i>Tweet10</i>	Aku ga keroso i min	Negatif

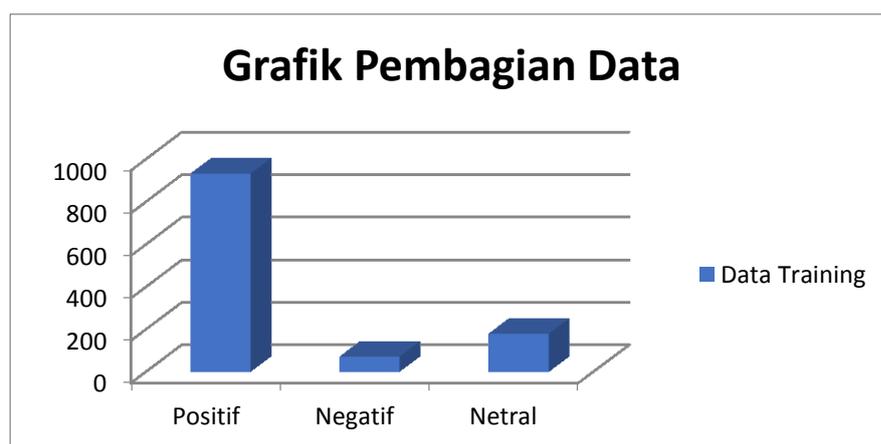
Tabel 7. Kasus Data *Training* Netral

<i>Tweet</i>	Kalimat	Label
<i>Tweet1</i>	Bmkg itu bakso malang karapitan group	Netral
<i>Tweet2</i>	Nggak ada yang dibuat rebahan	Netral
<i>Tweet3</i>	Kenceng apanya Grid? Akhir bulan aku pulang.	Netral
<i>Tweet4</i>	Aku itu punya aplikasi @infoBMKG tiap ada gempa selalu ada info, sekecil apapun	Netral
<i>Tweet5</i>	Gempi mksd e?	Netral
<i>Tweet6</i>	Mosok ee, pengalaman ledome wkw	Netral
<i>Tweet7</i>	Nyuci baju selesai dan hujan pun turun	Netral
<i>Tweet8</i>	Belum ada pak	Netral
<i>Tweet9</i>	Yo lek mbak e turu gk kerasa	Netral
<i>Tweet10</i>	Gk keroso min, turuku anteng paling, tangi2 wes adan subuh.	Netral
<i>Tweet4</i>	Aku itu punya aplikasi @infoBMKG tiap ada gempa selalu ada info, sekecil apapun	Netral
<i>Tweet5</i>	Gempi mksd e?	Netral
<i>Tweet6</i>	Mosok ee, pengalaman ledome wkw	Netral
<i>Tweet7</i>	Nyuci baju selesai dan hujan pun turun	Netral

<i>Tweet</i>	Kalimat	Label
<i>Tweet8</i>	Belum ada pak	Netral
<i>Tweet9</i>	Yo lek mbak e turu gk kerasa	Netral
<i>Tweet10</i>	Gk keroso min, turuku anteng paling, tangi2 wes adan subuh.	Netral
<i>Tweet4</i>	Aku itu punya aplikasi @infoBMKG tiap ada gempa selalu ada info, sekecil apapun	Netral
<i>Tweet5</i>	Gempi mksd e?	Netral
<i>Tweet6</i>	Mosok ee, pengalaman ledome wkw	Netral
<i>Tweet7</i>	Nyuci baju selesai dan hujan pun turun	Netral
<i>Tweet8</i>	Belum ada pak	Netral
<i>Tweet9</i>	Yo lek mbak e turu gk kerasa	Netral
<i>Tweet10</i>	Gk keroso min, turuku anteng paling, tangi2 wes adan subuh.	Netral

Tabel 8. Pembagian Data *Training*

No	<i>Class</i>	<i>Data Training</i>
1	Positif	931
2	Negatif	70
3	Netral	178
	Total	1179



Gambar 5. Grafik Pembagian Data

3.2.2 Akurasi Data

Akurasi adalah penghitungan dari perbandingan antara jumlah data dokumen yang relevan dan jumlah keseluruhan dokumen dalam database [15]. Menghitung akurasi dapat menggunakan persamaan 6.

$$Akurasi = \frac{Jumlah\ Dokumen\ Relevan}{Jumlah\ Keseluruhan\ Dokumen} \quad (6)$$

$$Akurasi = \frac{825}{1179}$$

$$Akurasi = 69,97\%$$

Berdasarkan hasil pembahasan, dari proses dan tahapan *preprocessing* yang dilakukan *Tweet* yang mengungkapkan sebuah ekspresi positif dan membuat *Tweet* tersebut terbaca negatif karena kesalahan. Hal ini dikarenakan masih ada yang terbaca negatif meskipun *Tweet* yang disertai menunjukkan *Tweet* positif. Tabel 9 menunjukkan contoh kesalahan *Tweet* yang didapat.

Tabel 9. Contoh Kesalahan *Tweet*

<i>Tweet</i>	Positif	Negatif	Netral
Bantu pantau hilal jodoh saya min @infoBMKG (positif)			√
Wahwaih (netral)		√	
Jalan ma aku aja (negatif)			√
Tidur min, mau temanin saya nonton bola? (negatif)	√		
min ini bener atau hoax y? (positif)		√	
Aku disemagati oleh diriku sendiri (positif)		√	
Notif dari tadi gak ada notif WA atau line isi nya notif (positif)		√	
Bahkan yang selalu serius pun butuh waktu rehat sejenak untuk menghela nafas, dan tersenyum (positif)		√	

Dengan menghilangkan salah satu tahapan *preprocessing* seperti yang penulis lakukan yaitu dengan menghilangkan tahapan Normalisasi fitur sangat berpengaruh dengan nilai akurasi yang dihasilkan karena dapat mengurangi hasil akurasi pada klasifikasi *Tweet*.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang dilakukan, maka terdapat beberapa kesimpulan yaitu :

1. Proses ekstraksi data dari *Twitter* BMKG Nasional menggunakan Bahasa pemrograman *Python* 3.74 dengan tahapan *Preprocessing* yang meliputi *casefolding*, *filtering*, *tokenisasi*, *slang replacement* dan *stopword removal*.
2. Sentimen analisis yang dilakukan adalah membuat tiga kelas yaitu positif, negatif dan netral menggunakan algoritma klasifikasi Naive Bayes.
3. Tingkat akurasi berdasarkan pengujian yang dilakukan adalah 68,97%.

REFERENCES

- [1] Apji, "Asosiasi Pengusaha Jasa Internet Indonesia," Jakarta, 2014.
- [2] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "DATA *TWITTER* KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *J. Ilm. Educ.*, vol. 7, no. 1, pp. 1–11, 2020.
- [3] A. Samad, H. Basari, B. Hussin, I. G. Pramudya, and J. Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *Procedia Eng.*, vol. 53, no. December, pp. 453–462, 2013, doi: 10.1016/j.proeng.2013.02.059.
- [4] M. A. Assuja and Saniati, "Analisis Sentimen *Tweet* Menggunakan," *J. Tek.*, vol. 10, no. 2, pp. 23–28, 2016.
- [5] A. P. Wijaya, H. A. Santoso, J. T. Informatika, U. Dian, and N. Semarang, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," *J. Appl. Intell. Syst.*, vol. 1, no. 1, pp. 48–55, 2016.
- [6] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 2, 2020.

- [7] A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, "Twitter sentiment analysis of movie reviews using machine learning technique," *Int. J. Eng. Technol.*, vol. 7, no. 6, pp. 2038–2044, 2016.
- [8] A. Surahman, "PENGEMBANGAN MARKET SEGMENTASI UNTUK MENCAPAI KEUNGGULAN BERSAING PADA E-MARKETPLACE," *J. Komput. dan Inform.*, vol. 15, no. 1, pp. 118–126, 2020.
- [9] B. R. Feldman, "Techniques and Applications for Sentiment Analysis," *Commun. Acm*, vol. 58, no. 4, pp. 83–89, 2013.
- [10] L. F. S. Coletta, N. F. F. da Silva, E. R. Hruschka, and E. R. H. Jr, "Combining classification and clustering for Tweet sentiment analysis Combining Classification and Clustering for Tweet Sentiment Analysis," in *Brazilian Conference on Intelligent Systems*, 2014, pp. 210–2015, doi: 10.1109/BRACIS.2014.46.
- [11] I. D. I, I. Amirulloh, and D. Rosiyadi, "Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda Naive Bayes," *J. Electr. Electron. Eng.*, vol. 1, no. 2, pp. 38–45, 2016.
- [12] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, pp. 86–91, 2020.
- [13] J. Ling and T. B. Oka, "ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE," *E-Jurnal Mat.*, vol. 3, no. 3, pp. 92–99, 2014.
- [14] A. Rahmawati, A. Marjuni, and J. Zeniarja, "Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksanaan Pilkada Serentak Menggunakan Algoritma Support Vector Machine," *CCIT J.*, vol. 10, no. 2, pp. 197–206, 2017.
- [15] A. Surahman, A. F. O. P. Pasaribu, and D. Darwis, "Ekstraksi Data Produk e-Marketplace Sebagai Strategi Pengolahan Segmentasi Pasar Menggunakan Web Crawler," *Sist. J. Sist. Inf.*, vol. 9, no. 1, pp. 73–81, 2020.