

ANALISIS SENTIMEN TERHADAP KONTEN YOUTUBE KRITIK SOSIAL DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER (NBC) DAN TF-IDF

Dustin Zaki Athariq^{1*}, Arief Wibowo²

^{1,2}Teknik Informatika, Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}2011501414@student.budiluhur.ac.id, ²arief.wibowo@budiluhur.ac.id

(* : corresponding author)

Abstrak-Media sosial telah menjadi bagian terpenting dalam kehidupan sehari-hari masyarakat Indonesia. Menurut data sebanyak 93,8% masyarakat Indonesia menggunakan media sosial YouTube secara rutin untuk berbagai keperluan, seperti mencari informasi, hiburan, dan berita terbaru. Salah satu konten yang sedang menarik perhatian adalah konten “Kuali Merah Putih” oleh Bobon Santoso, yang dimana menyoroti masalah kritik sosial terhadap kelaparan di Papua. Konten “Kuali Merah Putih” mendapatkan berbagai opini dari masyarakat Indonesia, yang dimana mencerminkan beragam sentimen positif, negatif atau netral. Untuk mengevaluasi tanggapan masyarakat terhadap konten akan dilakukan analisis sentimen. Penelitian bertujuan melihat hasil opini masyarakat terhadap konten “Kuali Merah Putih” yang dibuat oleh masyarakat. Metode *Naïve Bayes Classifier* (NBC) menunjukkan performa akurasi yang bagus dalam analisis sentimen komentar masyarakat mengenai konten “Kuali Merah Putih”, pada pelabelan berbasis aturan penulis dan divalidasi pakar mendapatkan akurasi sebesar 90%, sedangkan pelabelan *lexicon based* didapatkan akurasi sebesar 84%. Hasil dari kedua pelabelan menunjukkan masyarakat memberikan komentar positif dan mendukung terhadap konten “Kuali Merah Putih”. pada pelabelan persepsi peneliti dan divalidasi pakar didapatkan 250 berkomentar positif, sedangkan *lexicon based* didapatkan sebanyak 250 komentar positif. Persepsi dan sentimen publik terhadap isu kritik sosial kelaparan di Papua yang terungkap melalui konten “Kuali Merah Putih” sangat positif. Temuan ini menunjukkan bahwa masyarakat Indonesia memiliki kepedulian yang tinggi terhadap permasalahan sosial di tanah air dan mengharagai upaya-upaya untuk menyuarakan isu-isu tersebut.

Kata Kunci: YouTube, Papua, Kritik Sosial, TF-IDF, *Sentimen Analysis, Naive Bayes Classifier*.

SENTIMENT ANALYSIS OF YOUTUBE SOCIAL CRITICAL CONTENT USING NAÏVE BAYES CLASSIFIER (NBC) AND TF-IDF ALGORITHMS

Abstract-*Social Media has become the most important part of everyday life of Indonesian people. According to data, 93.8% of Indonesians regularly use YouTube for various purposes, such as searching for information, entertainment, and the latest news. One of the content that's attracting attention is Bobon Santoso's "White Red Quali" content, which highlights the social criticism of famine in Papua. The contents of "White Red Quali" gained various opinions from Indonesian society, which reflected a variety of positive, negative or neutral sentiments. To evaluate the public's response to content, sentimental analysis will be carried out. The research is aimed at looking at the public opinion of the content of the "White Red Tree" created by the public. The Naïve Bayes Classifier method (NBC) showed good accuracy performance in the analysis of public comments sentiment on the content of "Red White Quali", on authors' rule-based labelling and expert validation obtaining accuracy of 90%, while lexicon based labelling obtained accurations of 84%. Results from both labelling showed public giving positive comments and support for the "Red white Quali" content. On the researchers' perception labeling and validation of experts obtains 250 positive comments, whereas lexicon-based obtained as many as 250 negative comments. The perception and public sentiment of the issue of social criticism of famine in Papua that is expressed through the content of "White Red Quali" is very positive. The findings indicate that Indonesians have a high concern for social issues in their homeland and appreciate efforts to raise them.*

Keywords: YouTube, Papua, Kritik Sosial, TF-IDF, *Sentimen Analysis, Naive Bayes Classifier*

1. PENDAHULUAN

Indonesia saat ini mengalami kemajuan pesat di bidang teknologi informasi digital. Media sosial digunakan masyarakat Indonesia sehari-hari untuk mendapatkan informasi dengan cepat dan mudah. YouTube, sebagai salah satu media sosial, memungkinkan akses cepat ke berbagai sumber tanpa batas geografis. Masyarakat menggunakan YouTube untuk berbagai keperluan, seperti melihat berita, mendengarkan lagu, menonton video, dan mencari informasi terkini. Salah satu konten yang populer adalah “Kuali Merah Putih” oleh Bobon Santoso, yang mengkritik isu kelaparan di Papua.

Papua adalah provinsi paling timur di Indonesia dengan kondisi geografis yang menantang, mengakibatkan bahan pokok menjadi mahal dan menyebabkan kelaparan. Pemerintah daerah Yahukimo menetapkan status tanggap darurat bencana pada 21 Oktober 2023 hingga 1 November 2023 setelah 23 warga meninggal karena kelaparan. Ini bukan pertama kalinya kasus kelaparan terjadi di Papua, dengan laporan kematian lainnya pada bulan Agustus 2023, juga dilaporkan beberapa kematian akibat krisis pangan.

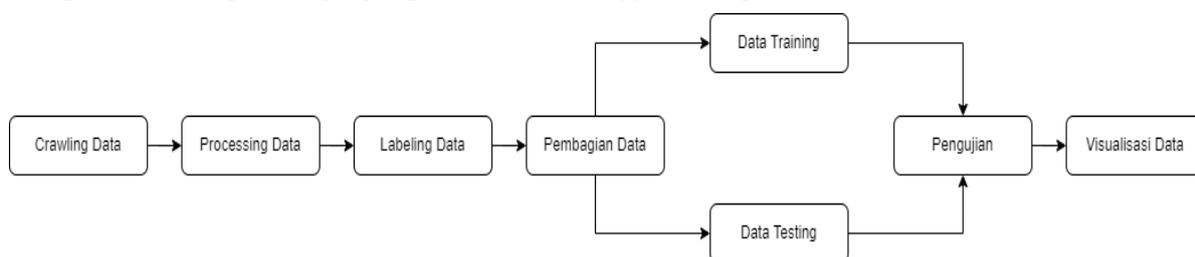
Konten “Kuali Merah Putih” yang dibuat oleh Bobon Santoso memicu berbagai tanggapan dari masyarakat, termasuk opini positif, negatif, dan netral. Analisis menjadi penting untuk mengklasifikasikan opini-opini, guna memahami lebih dalam bagaimana publik memberikan reaksi terhadap konten yang menyentuh isu sensitif kelaparan di Papua. Dalam skala besar, pengklasifikasian konvensional akan sangat tidak efisien, sehingga peran analisis sentimen sangat diperlukan untuk mempercepat proses klasifikasi komentar. Penelitian terkait analisis sentimen di media sosial telah dilakukan dengan berbagai metode. Misalnya, sebuah studi sebelumnya menggunakan metode *Naïve Bayes* untuk menganalisis sentimen terkait Pinjaman Online di media sosial Twitter. Studi tersebut menunjukkan bahwa *Naïve Bayes*, setelah proses *text preprocessing* dan pembobotan menggunakan metode TF-IDF, menghasilkan akurasi sebesar 80%. Dari 2912 data yang tersedia, 68,61% adalah sentimen negatif, sementara 31,39% adalah sentimen positif [1]. Selain itu, penelitian lain yang menggunakan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) untuk analisis sentimen aplikasi Gojek juga menunjukkan hasil yang berbeda, dengan K-NN mendapatkan akurasi sebesar 82,14% dan SVM mencapai akurasi 87,98% [2].

Namun, penelitian-penelitian sebelumnya berfokus pada topik dan platform yang berbeda seperti pinjaman online di Twitter dan aplikasi Gojek. Sementara itu, studi yang spesifik pada analisis sentimen terkait kritik sosial di YouTube, khususnya pada isu kelaparan di Papua, belum banyak diteliti. Selain itu, meskipun metode *Naïve Bayes Classifier* (NBC) dan TF-IDF telah diterapkan pada platform lain, penggunaan metode ini dalam kritik sosial di platform YouTube, terutama untuk menganalisis reaksi terhadap konten seperti “Kuali Merah Putih” oleh Bobon Santoso, masih belum dieksplorasi secara mendalam. Analisis sentimen di YouTube menghadirkan tantangan unik, mengingat sifat platform yang memungkinkan reaksi yang lebih beragam dan lebih luas dibandingkan dengan media sosial sosial seperti Twitter.

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan melakukan analisis sentimen terhadap 5,998 komentar dari *channel* YouTube Bobon Santoso terkait konten “Kuali Merah Putih” yang mengkritik isu kelaparan di Papua. Metode *Naïve Bayes Classifier* (NBC) dan TF-IDF akan diterapkan untuk mengklasifikasikan sentimen publik terhadap konten kritik sosial ini. Dengan fokus pada platform YouTube, penelitian ini akan memberikan wawasan baru tentang algoritma yang sama dapat berfungsi dalam konteks yang berbeda dan menawarkan pandangan yang lebih dalam tentang respon publik terhadap isu sosial di Indonesia.

2. METODE PENELITIAN

Dalam membangun aplikasi analisis sentimen penelitian ini menggunakan implementasi algoritma *Naïve Bayes Classifier* (NBC). Dimana sistem memiliki beberapa tahapan yang akan dilakukan. Gambar 1 di bawah ini merupakan alur tahapan dari pengumpulan data awal hingga menampilkan hasil:



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Pada penelitian ini data didapatkan dengan melakukan *crawling* komentar pada media sosial YouTube dengan menggunakan bahasa pemrograman Python terhadap konten Bobon Santoso dengan judul “Kuali Merah Putih”. Data komentar YouTube didapatkan sebanyak 5,998 komentar dari total 6 video “Kuali Merah Putih”. Pengambilan data dilakukan pada tanggal 03 Maret 2024 dan 30 April 2024, dan memiliki 4 atribut antara lain *username*, *id_YouTube*, *time* dan *text*.

2.2 Preprocessing

Preprocessing adalah sebuah tahapan untuk melakukan pembersihan dari kata-kata atau komentar yang tidak mengandung makna di dalamnya [3]. Dalam proses *processing* ini akan melakukan beberapa tahapan yang dimana melakukan *filtering* data, konversi emotikon, *casefolding*, *tokenizing*, *cleansing*, normalisasi kata, *stopword*, *stemming*, dan *drop duplicate*.

- Filtering* yaitu suatu proses dimana data diambil berdasarkan kata kunci yang diberikan. *Filtering* data ini dilakukan untuk memastikan bahwa hanya mengambil data yang relevan untuk di proses lebih lanjut, sehingga nantinya analisis dilakukan menjadi lebih efektif dan efisien.
- Konversi emotikon merupakan salah satu cara untuk mengekspresikan ekspresi perasaan secara tekstual [4].
- Casefolding* adalah sebuah tahapan untuk mengubah data di dalam sebuah dokumen yang memiliki huruf kapital akan dirubah menjadi huruf kecil dari a hingga z [5].
- Tokenizing* adalah sebuah metode untuk memisahkan kata dari dalam suatu kalimat dengan tujuan untuk mempermudah dalam melakukan processing data.
- Cleansing* adalah suatu tahapan untuk melakukan pembersihan pada teks yang dimana di dalamnya menghapus simbol dan karakter seperti *url* (<https://>), *hashtag* (#), *mentions*(@), angka, tanda baca, dan menghapus spasi berlebihan.
- Normalisasi teks adalah sebuah proses untuk merubah kata yang tidak sesuai dengan ejaan bahasa Indonesia yang baku(EYD) bagus berupa kata singkatan, gaul atau modern. atau kesalahan dala penulisan lainnya [6].
- Stopword* adalah sekumpulan kata-kata umum yang sering digunakan dalam teks tetapi dianggap tidak memiliki makna signifikan dalam analisis sentimen.
- Stemming* adalah suatu tahapan untuk mengubah kata menjadi kata dasar menurut kaidah bahasa Indonesia yang benar.
- Drop duplicate* adalah sebuah tahapan untuk menghapus data yang lebih dari satu atau data duplikat. Biasanya, ketika melakukan *crawling*, data yang diperoleh bisa tersimpan lebih dari satu. Oleh karena itu tahapan ini digunakan untuk menghapus data yang duplikat.

2.3 Labeling Data

Labeling adalah sebuah proses untuk melakukan pengelompokan data menjadi 3 bagian yaitu positif, negatif atau netral berdasarkan skor nilai yang dihasilkan dari pelabelan berbasis aturan dengan persepsi peneliti dan *lexicon based* dengan mengimplementasikan kamus kata positif dan negatif dari fajri91 [7].

2.4 Split Data

Split data adalah sebuah proses untuk membagi data menjadi data *training* dan *testing*. Pada penelitian yang dilakukan oleh [8], *split* data merupakan langkah penting yang dapat mempengaruhi akurasi pengolahan data. Dalam penelitian ini, data akan dibagi menjadi data *training* dan data *testing* dengan rasio 70:30.

2.5 TF-IDF

TF-IDF adalah suatu kata term yang ada disetiap dokumen untuk diberikan suatu nilai bobot. Metode ini menggabungkan 2 tahapan yang dimana menghitung jumlah ke munculan sebuah kata di dalam suatu *document* yang disebut dengan *term frequency* (TF) dan *inverse document frequency* (IDF) [9]. Berikut rumus yang digunakan dalam TF-IDF:

- Term Frequency* (TF)

$$TF(term) = \frac{\text{jumlah kemunculan term dalam document}}{\text{total jumlah term dalam document}} \quad (2.1)$$

- Inverse Document Frequency*

$$idf(term) = \text{Log} \left(\frac{N}{DF(term)} \right) + 1 \quad (2.2)$$

c. TF-IDF

$$tfidf(term) = tf(term) * idf(term) \quad (2.3)$$

| | | |
|----------------|---|--|
| Keterangan | : | |
| $TF(term)$ | : | Jumlah ke munculan kata ($term$) di dalam dokumen tersebut |
| T | : | Jumlah total ke munculan kata di dalam dokumen tersebut |
| N | : | Jumlah keseluruhan dokumen |
| $DF(term)$ | : | Jumlah ke munculan kata di seluruh dokumen |
| $tf-idf(term)$ | : | Bobot kata ($term$) terhadap dokumen |

2.6 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) merupakan metode dengan klasifikasi yang sesuai dengan teorema Bayes. pada tahapan ini akan melakukan pengklasifikasian berdasarkan *probabilitas* yang bisa terjadi dimasa depan [10]. Berikut rumus umum *Naïve Bayes Classifier*(NBC):

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (2.4)$$

| | | |
|------------|---|--|
| Keterangan | : | |
| X | = | Data dengan kelas belum diketahui |
| C | = | Hipotesa data C merupakan suatu kelas spesifik |
| $P(X C)$ | = | <i>Probabilitas</i> hipotesis C berdasarkan kondisi postior <i>probability</i>) |
| $P(C)$ | = | <i>Probabilitas</i> hipotesis C (<i>prior probability</i>) |

Pada suatu *dataset*, pemilihan memungkinkan adanya nilai nol pada model *probabilitas*. Nilai nol tersebut dapat menyebabkan *Naïve Bayes Classifier* (NBC) tidak dapat melakukan klasifikasi, maka dari itu diperlukan metode *one smoothing* untuk menjadi solusi dan menghindari angka nol, karena dalam perhitungan nilai n akan ditambah 1. Berikut rumus yang digunakan dalam metode *one smoothing*:

$$P(F_k | C_1) = \frac{n_k + 1}{n + vocabulary} \quad (2.6)$$

| | | |
|-------------------|---|--|
| $P(F_k C_i)$ | = | Nilai <i>probabilitas</i> F_k pada kelas C_i |
| n_k | = | Jumlah data atribut n_k |
| N | = | Jumlah data pada kelas dari atribut n_k |
| <i>vocabulary</i> | = | Jumlah data dari dalam class n_k |

2.7 Confusion Matrix

Confusion matrix adalah sebuah metode yang digunakan untuk melakukan evaluasi kinerja program sari suatu model klasifikasi. *Confusion matrix* digunakan sebagai evaluasi klasifikasi yang diharapkan dapat mengetahui seberapa akurat metode yang di evaluasi [11].

3. HASIL DAN PEMBAHASAN

Hasil dari implementasi pada metode penelitian ini terdapat beberapa tahapan dengan penjabaran sebagai berikut:

3.1 Pengumpulan Data (*Crawling*)

Dalam tahap pengambilan data, penelitian ini menggunakan API *library* yaitu YouTube data API v3 yang diperoleh dari Google *Developers*. Pengumpulan ini menggunakan sentimen komentar video YouTube pada tanggal 03 Maret 2024 dan 30 April 2024 didapatkan data sebanyak 5,998 komentar dari konten “Kualiti Merah Putih”. Berikut ini contoh data hasil pengumpulan data *crawling* yang dapat di lihat pada Tabel 1.

Tabel 1. Pengumpulan Data

| No | Username | Time | Text |
|----|-----------------------|--------------------------|--|
| 1 | @AlamPedagangkeliling | April 3, 2024, 5:58 a.m. | Makasih bang bobon smoga rejeki nya lancar terus |
| 2 | @cahyaofficial7970 | April 3, 2024, 5:47 a.m. | Beda banget tatapan mata mereka kelihatan tulus banget, di sini rata2 pada malak2 gak jelas mabok2 |

| | | | |
|-------|-------------------|--------------------------|---|
| 3 | @andrewtv7863 | April 3, 2024, 3:41 a.m. | pinggir jalan, kalau rusuh satu kampung di bawa. Perantau jadi imbasnya Terenyuh, ternyata sodara" kita sangat bagus" ramah, salam dari jawa tengah untuk saudara"ku di Papua.. 🗣️ |
| ... | ... | ... | ... |
| 5,998 | @sudirohusodo8027 | May 3, 2024, 3:11 a.m | banyak yg usul gaji dan pasilitas TNI ditingkatkan namun apakah tidak rentan di korupsi kah |

3.2 Preprocessing

Tahap selanjutnya yaitu melakukan *preprocessing*. Tahap ini dilakukan untuk membersihkan data yang tidak terstruktur menjadi data lebih terstruktur dan bersih. Tahap *preprocessing* terdiri dari sembilan proses, yaitu *filtering* data, konversi emotikon, *casefolding*, *tokenizing*, *cleansing*, normalisasi kata, *stopword*, *stemming*, dan *drop duplicate*.

a. Filtering

Proses *filtering* yaitu pembersihan data berdasarkan *keyword* yang telah di tentukan yaitu Papua dan Merauke. Data sebelum dilakukan *filtering* berjumlah 5,998 data dan setelah dilakukan *filtering* didapatkan data sebanyak 836 data.

b. Konversi Emoticon

Proses konversi emotikon yaitu merubah bentuk emotikon menjadi teks untuk diikutkan dalam proses analisa. Proses konversi emotikon dapat di lihat pada tabel 2.

Tabel 2. Konvert Emotikon

| Text | Konversi |
|--|--|
| "KITA TIDAK AKAN PERNAH MENINGGALKAN PAPUA SMP KPN PUN" @bobon santoso 🗣️ | "KITA TIDAK AKAN PERNAH MENINGGALKAN PAPUA SMP KPN PUN" @bobon santoso bangga |

c. Case Folding

Proses *case folding* yaitu mengubah kata dari huruf kapital menjadi huruf kecil dimana membuat data menjadi terstruktur. Proses *case folding* dapat di lihat pada tabel 3.

Tabel 3 Case Folding

| Konversi | Case Folding |
|--|---|
| "KITA TIDAK AKAN PERNAH MENINGGALKAN PAPUA SMP KPN PUN" @bobon santoso bangga | "kita tidak akan pernah meninggalkan papua smp kpn pun" @bobon santoso bangga |

d. Tokenizing

Proses *tokenizing* yaitu memecahkan sebuah kalimat menjadi bentuk kata-kata yang berfungsi untuk memudahkan dalam membersihkan sebuah teks. Proses *tokenizing* dapat di lihat pada tabel 4.

Tabel 4. Tokenizing

| Case Folding | Tokenizing |
|---|--|
| "kita tidak akan pernah meninggalkan papua smp kpn pun" @bobon santoso bangga | ['"kita', ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['smp'], ['kpn'], ['pun"'], [' @bobon santoso'], ['bangga'] |

e. Cleansing

Proses *cleansing* yaitu menghapus data yang tidak digunakan seperti *link url*, *mentions*, angka, spasi berlebih dan karakter khusus. Proses *cleansing* dapat di lihat pada tabel 5.

Tabel 5. Cleansing

| <i>Tokenizing</i> | <i>Cleansing</i> |
|---|--|
| ['"kita'], ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['smp'], ['kpn'], ['pun"'], [' @bobon santoso'], ['bangga'] | ['kita'], ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['smp'], ['kpn'], ['pun'], ['bobon santoso'], ['bangga'] |

f. Normalisasi Teks

Proses normalisasi teks yaitu menggantikan kata-kata alay menjadi ejaan yang sesuai dengan EYD. Proses normalisasi teks dapat di lihat pada tabel 6.

Tabel 6. Normalisasi Teks

| <i>Cleansing</i> | Normalisasi Teks |
|--|---|
| ['kita'], ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['smp'], ['kpn'], ['pun'], ['bobon santoso'], ['bangga'] | ['kita'], ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['sampai'], ['kapan'], ['pun'], ['bobon santoso'], ['bangga'] |

g. Stopword

Proses *stopword* yaitu menghapus teks yang tidak digunakan dalam proses analisa biasanya berupa kata penghubung seperti yang, tidak, dan sebagainya. Proses *stopword* dapat di lihat pada tabel 7.

Tabel 7. Stopword

| Normalisasi Teks | <i>Stopword</i> |
|---|---|
| ['kita'], ['tidak'], ['akan'], ['pernah'], ['meninggalkan'], ['papua'], ['sampai'], ['kapan'], ['pun'], ['bobon santoso'], ['bangga'] | ['meninggalkan'], ['papua'], ['bobon santoso'], ['bangga'] |

h. Stemming

Proses *stemming* yaitu menghapus kata imbuhan dan mengembalikan kata menjadi kata dasarnya. Proses *stemming* dapat di lihat pada tabel 8.

Tabel 8. Stemming

| <i>Stopword</i> | <i>Stemming</i> |
|---|------------------------------------|
| ['meninggalkan'], ['papua'], ['bobon santoso'], ['bangga'] | tinggal papua bobon santoso bangga |

i. Drop Duplicate

Proses *drop duplicate* yaitu menghapus data yang memiliki teks atau komentar yang sama.

3.3 Labeling Data

Proses *labeling* yaitu sebuah tahap untuk melakukan pemberian label positif, negatif atau netral. Dalam hal ini label positif menunjukkan dukungan terhadap konten, label negatif menunjukkan ketidaksetujuan, dan label netral menunjukkan netralitas terhadap konten. Proses *labeling* dapat di lihat pada tabel 9.

Tabel 9. Labeling Data

| Teks | Label Konvensional | Label Kamus |
|---|--------------------|-------------|
| tinggal Papua bobon santoso | positif | netral |
| salut bobon santoso mending masak teman Papua ketimbang rohingnya Indonesia lapar | negatif | negatif |
| konten tunggu sehat abang bobon sehat warga Papua doa | positif | positif |

3.4 Split Data

Pada proses ini data yang telah dilakukan *labeling* akan dilakukan pembagian sebanyak 70% *training* dan 30% *testing* dari dua kelas yaitu pelabelan berbasis aturan dengan persepsi peneliti yang telah divalidasi oleh pakar di bidang Ilmu Politik dan *lexicon based*. Proses *split data* dapat di lihat pada tabel 10.

Tabel 10. Split Data

| Dataset | Jumlah |
|---------------|--------|
| Data Training | 585 |
| Data Testing | 251 |
| Total Data | 836 |

3.5 TF-IDF

Langkah selanjutnya adalah memberikan bobot ke setiap kata yang ada di dalam dokumen. proses ini melibatkan perhitungan *term frequency* (TF), *inverse document frequency* (IDF) dan TF-IDF. Untuk ilustrasi perhitungan akan menggunakan 1 data dari total 585 data *training*. Proses untuk mencari nilai *term frequency* (TF) dapat di lihat pada tabel 11.

Tabel 11. Term Frequency (TF)

| Stemming | d1 | d2 | d3 |
|----------|-------|------|-------|
| bingung | 0.167 | 0 | 0 |
| bawa | 0.167 | 0 | 0 |
| kuali | 0.167 | 0 | 0 |
| papua | 0.167 | 0.25 | 0.143 |
| bang | 0.167 | 0 | 0 |
| mohon | 0.167 | 0 | 0 |
| orang | 0 | 0.25 | 0.143 |
| hidung | 0 | 0.25 | 0 |
| mancung | 0 | 0.25 | 0 |
| suka | 0 | 0 | 0.143 |
| senang | 0 | 0 | 0.143 |
| kalo | 0 | 0 | 0.143 |
| lihat | 0 | 0 | 0.143 |
| sopan | 0 | 0 | 0.143 |

Selanjutnya yaitu mencari nilai *inverse document frequency* (IDF). Proses untuk mencari nilai *inverse document frequency* (IDF) dapat di lihat pada tabel 12.

Tabel 12. Inverse Document Frequency (IDF)

| Stemming | ke munculan | Rumus | IDF |
|----------|-------------|----------------------------|-------|
| bingung | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| bawa | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| kuali | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| papua | 3 | $\text{Log}_{10}(3/1+3)+1$ | 0.875 |
| bang | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| mohon | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| orang | 2 | $\text{Log}_{10}(3/1+2)+1$ | 1 |
| hidung | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| mancung | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| suka | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| senang | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| kalo | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| lihat | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |
| sopan | 1 | $\text{Log}_{10}(3/1+1)+1$ | 1.176 |

Setelah didapatkan *term frequency* (TF) dan *inverse document frequency* (IDF) selanjutnya menghitung nilai TF-IDF. Dimana Proses untuk mencari nilai TF-IDF dapat di lihat pada tabel 13.

Tabel 13. Menghitung TF * IDF

| Stemming | d1 | d2 | d3 |
|----------|-------|-------|-------|
| bingung | 0.196 | 0 | 0 |
| bawa | 0.196 | 0 | 0 |
| kuali | 0.196 | 0 | 0 |
| papua | 0.146 | 0.219 | 0.125 |
| bang | 0.196 | 0 | 0 |
| mohon | 0.196 | 0 | 0 |

| | | | |
|---------|-------|-------|-------|
| orang | 0 | 0.25 | 0.143 |
| hidung | 0 | 0.294 | 0 |
| mancung | 0 | 0.294 | 0 |
| suka | 0 | 0 | 0.168 |
| senang | 0 | 0 | 0.168 |
| kalo | 0 | 0 | 0.168 |
| lihat | 0 | 0 | 0.168 |
| sopan | 0 | 0 | 0.168 |
| Jumlah | 1.126 | 1.057 | 1.108 |

Dengan perhitungan di atas maka didapatkan hasil nilai TF, IDF dan TF-IDF. Dengan proses ini dapat disimpulkan bahwa seluruh data yang ada dilakukan pembobotan dengan proses yang sama.

3.6 Naïve Bayes Classifier

Setelah dilakukan pembobotan kata selanjutnya akan masuk tahap klasifikasi data dengan menggunakan algoritma *Naïve Bayes Classifier* (NBC). Tahap ini dilakukan pemodelan dan mengklasifikasikan data *testing* terhadap data yang telah dilakukan *training*. Dalam proses ini akan melakukan perbandingan antara hasil pelabelan berbasis aturan dengan persepsi peneliti yang telah divalidasi oleh pakar di bidang Ilmu Politik dan *lexicon based*. Untuk ilustrasi dalam klasifikasi akan menggunakan 2 data *testing* dari total sebanyak 251 komentar. Berikut proses perhitungannya.

a. Menghitung Prior Class

Langkah pertama yaitu menghitung nilai *prior class*. Proses untuk menghitung nilai *prior class* dapat dilihat pada tabel 14.

Tabel 14. Menghitung Skor Prior Class

| Kelas | Jumlah | Hasil |
|---------|--------|-------|
| netral | 1/3 | 0.333 |
| positif | 1/3 | 0.333 |
| negatif | 1/3 | 0.333 |

b. Prediksi Pelabelan Berbasis Aturan Penulis Dan *Lexicon Based*

1. Prediksi data *testing* “lihat masyarakat papua moga perintah perhati pikat”

Langkah pertama adalah mencari *likelihood* untuk menentukan $P(X|C)$. Proses untuk mencari *likelihood* dalam menentukan $P(X|C)$ dapat dilihat pada tabel 15.

Tabel 15. Menghitung *Likelihood* $P(X|C)$ Netral, Negatif dan Positif

| Kata | Kata Unik | Menghitung | Hasil perhitungan Netral |
|------------|-----------|--|---------------------------|
| lihat | 0 | $\text{Log}(0 + 1) / (1.126 + 14)$ | $\text{Log}(0.666)$ |
| masyarakat | 0 | $\text{Log}(0 + 1) / (1.126 + 14)$ | $\text{Log}(0.666)$ |
| papua | 0.146 | $\text{Log}(0.146 + 1) / (1.126 + 14)$ | $\text{Log}(0.076)$ |
| moga | 0 | $\text{Log}(0 + 1) / (1.126 + 14)$ | $\text{Log}(0.666)$ |
| perhati | 0 | $\text{Log}(0 + 1) / (1.126 + 14)$ | $\text{Log}(0.666)$ |
| pikat | 0 | $\text{Log}(0 + 1) / (1.126 + 14)$ | $\text{Log}(0.666)$ |
| Kata | Kata Unik | Menghitung | Hasil perhitungan Negatif |
| lihat | 0 | $\text{Log}(0 + 1) / (1.057 + 14)$ | $\text{Log}(0.666)$ |
| masyarakat | 0 | $\text{Log}(0 + 1) / (1.057 + 14)$ | $\text{Log}(0.666)$ |
| papua | 0.219 | $\text{Log}(0.219 + 1) / (1.057 + 14)$ | $\text{Log}(0.081)$ |
| moga | 0 | $\text{Log}(0 + 1) / (1.057 + 14)$ | $\text{Log}(0.666)$ |
| perhati | 0 | $\text{Log}(0 + 1) / (1.057 + 14)$ | $\text{Log}(0.666)$ |
| pikat | 0 | $\text{Log}(0 + 1) / (1.057 + 14)$ | $\text{Log}(0.666)$ |
| Kata | Kata Unik | Menghitung | Hasil perhitungan Positif |
| lihat | 0 | $\text{Log}(0 + 1) / (1.108 + 14)$ | $\text{Log}(0.666)$ |
| masyarakat | 0 | $\text{Log}(0 + 1) / (1.108 + 14)$ | $\text{Log}(0.666)$ |
| papua | 0.125 | $\text{Log}(0.125 + 1) / (1.108 + 14)$ | $\text{Log}(0.074)$ |
| moga | 0 | $\text{Log}(0 + 1) / (1.108 + 14)$ | $\text{Log}(0.666)$ |
| perhati | 0 | $\text{Log}(0 + 1) / (1.108 + 14)$ | $\text{Log}(0.666)$ |
| pikat | 0 | $\text{Log}(0 + 1) / (1.108 + 14)$ | $\text{Log}(0.666)$ |

Setelah didapatkan *likelihood* untuk kata yang muncul di setiap kelas, selanjutnya adalah tahap menghitung *skor log-probabilitas* nilai di setiap kelas untuk mendapatkan prediksi kelasnya. Proses untuk melakukan perhitungan *skor log-probabilitas* dapat di lihat pada tabel 16.

Table 16. Menghitung *Skor Log-Probabilitas* Netral, Negatif dan Positif

| Keterangan | Hasil |
|-----------------------------|---|
| $Log(P(\text{Netral}))$ | = 0.333 |
| $Log P(X \text{netral})$ | = $0.666 * 0.666 * 0.076 * 0.666 * 0.666 * 0.666 * 0.666$ = $0.333 * 0.666 * 0.666 * 0.076 * 0.666 * 0.666 * 0.666 * 0.666$ = 2.1083929501762615e-09 |
| Keterangan | Hasil |
| $Log(P(\text{Negatif}))$ | = 0.333 |
| $Log P(X \text{negatif})$ | = $0.666 * 0.666 * 0.076 * 0.666 * 0.666 * 0.666 * 0.666$ = $0.333 * 0.666 * 0.666 * 0.666 * 0.081 * 0.666 * 0.666 * 0.666 * 0.666$ = 2.315622975092322e-09 |
| $Log(P(\text{Positif}))$ | = 0.333 |
| $Log P(X \text{positif})$ | = $0.666 * 0.666 * 0.076 * 0.666 * 0.666 * 0.666 * 0.666$ = $0.333 * 0.666 * 0.666 * 0.074 * 0.666 * 0.666 * 0.666 * 0.666$ = 2.4380767432297803e-09 |

Dengan ini maka didapatkan hasil prediksi dari pelabelan berbasis aturan penulis yang telah divalidasi oleh pakar di bidang Ilmu Politik dan *lexicon based* terhadap “lihat masyarakat papua moga perintah perhati pikat” adalah positif. Pada hasil klasifikasi diatas dengan menggunakan 1 data *testing* sebagai contoh perhitungan dapat disimpulkan bahwa iterasi terhadap seluruh data *testing* lainnya dapat dilakukan dengan proses pengujian yang sama

3.7 Pengujian *Confusion Matrix*

Selanjutnya adalah pengujian dengan *confusion matrix* yang dimana bagian yang penting dari setiap fase pengembangan sistem. Dalam hal ini untuk melakukan evaluasi, menganalisis dan mengetahui hasil tingkat akurasi yang diperoleh dari keseluruhan data uji sebanyak 251 data dari hasil prediksi algoritma *Naïve Bayes Classifier* (NBC). Pengujian ini ditampilkan dalam bentuk table *confusion*, hasil akurasi, *precision*, *recall*, dan *f1-score*. Gambar 2 di bawah ini menampilkan hasil *confusion matrix* secara konvensional dan *lexicon based* (kamus).

| Confusion Matrix Manual | | | | Confusion Matrix Kamus | | | |
|-------------------------|---------|---------|--------|------------------------|---------|---------|--------|
| Actual / Predicted | Positif | Negatif | Netral | Actual / Predicted | Positif | Negatif | Netral |
| positif | 226 | 0 | 1 | positif | 212 | 0 | 1 |
| negatif | 10 | 0 | 0 | negatif | 26 | 0 | 0 |
| netral | 14 | 0 | 0 | netral | 12 | 0 | 0 |

Gambar 2. *Confusion Matrix* Convensional dan *Lexicon Based*

Selanjutnya adalah melakukan pengujian berdasarkan kelasnya, dimana akan menghitung masing-masing kelas *precision*, *recall* dan *f1-Score* (positif, negatif dan netral). Selain itu, juga akan menghitung nilai akurasi yang didapatkan dalam tahap ini, perhitungan dilakukan berdasarkan hasil label antara pelabelan berbasis aturan penulis dengan persepsi peneliti divalidasi oleh pakar dan *lexicon based*. Berikut ini Gambar 3 menampilkan hasil evaluasi *matrix*:

| Metrics Manual | | | | | Metrics Kamus | | | | |
|-----------------|-----------|--------|----------|---------|-----------------|-----------|--------|----------|---------|
| Class | Precision | Recall | F1 Score | Support | Class | Precision | Recall | F1 Score | Support |
| positif | 0.90 | 1.00 | 0.95 | 227 | positif | 0.85 | 1.00 | 0.92 | 213 |
| negatif | 0.00 | 0.00 | 0.00 | 10 | negatif | 0.00 | 0.00 | 0.00 | 26 |
| netral | 0.00 | 0.00 | 0.00 | 14 | netral | 0.00 | 0.00 | 0.00 | 12 |
| Accuracy | 0.90 | | | | Accuracy | 0.84 | | | |

Gambar 3. Evaluasi *Matrix* Konvensional dan *lexicon based*

Setelah dilakukan pengujian terhadap pelabelan dengan persepsi peneliti dan divalidasi pakar didapatkan akurasi sebesar 90% dengan jumlah hasil prediksi komentar sebanyak 250 positif, sedangkan terhadap pelabelan *lexicon based* didapatkan akurasi sebesar 85% dengan jumlah hasil prediksi komentar sebanyak 250 positif.

4. KESIMPULAN

Metode *Naïve Bayes Classifier* (NBC) menunjukkan performa akurasi yang bagus dalam analisis sentimen komentar masyarakat mengenai konten “Kuali Merah Putih”, pada pelabelan berbasis aturan penulis dan divalidasi pakar mendapatkan akurasi sebesar 90%, sedangkan pelabelan *lexicon based* didapatkan akurasi sebesar 84%. Kedua pelabelan mengindikasikan bahwa masyarakat memberikan komentar positif dan mendukung terhadap konten tersebut dengan 250 komentar positif teridentifikasi pada kedua metode pelabelan. Persepsi dan sentimen publik terhadap isu kritik sosial kelaparan di Papua yang terungkap melalui konten “Kuali Merah Putih” sangat positif. Mayoritas komentar masyarakat menunjukkan dukungan dan apresiasi terhadap upaya konten tersebut dalam mengangkat isu tersebut. Komentar-komentar positif tersebut mencerminkan adanya kesadaran dan kepedulian publik terhadap masalah kelaparan Papua, serta harapan bahwa konten “Kuali Merah Putih” dapat memberikan dampak positif dalam mengatasi masalah tersebut. Hal ini menunjukkan bahwa konten tersebut berhasil menyentuh hati masyarakat dan membangkitkan semangat untuk membantu saudara-saudara di Papua yang mengalami kelaparan.

DAFTAR PUSTAKA

- [1] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, “Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 1340–1348, 2023, doi: 10.30865/klik.v3i6.936.
- [2] M. N. Muttaqin and I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [3] A. A. Ningtyas, A. Solichin, and R. Pradana, “Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Naïve Bayes,” *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 20, no. 1, p. 9, 2023, doi: 10.36080/bit.v20i1.2317.
- [4] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, “Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 207–215, Apr. 2023, doi: 10.37859/coscitech.v4i1.4790.
- [5] N. Chatrina Siregar, R. Ruli, A. Siregar, ; M Yoga, and D. Sudirman, “Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ),” 2020.
- [6] V. R. A. Sejati, Painem, Ferdiansyah, and W. Pramusinto, “Pada Google Play Store Dengan Multinomial Naïve Sentiment Analysis of Ipusnas Application User Reviews on Google Play Store With Multinomial Naïve,” *SENAFTI Semin. Nas. Mahasiswa Fak. Teknol. Inf.*, vol. 2, no. 1, pp. 182–190, 2023.
- [7] M. I. Ahmadi, D. Gustian, and F. Sembiring, “Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes,” 2021.
- [8] Baiq Nurul Azmi, Arief Hermawan, and Donny Avianto, “Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.
- [9] H. Sari, G. Leonarde Ginting, and T. Zebua, “Penerapan Algoritma Text Mining dan TF-IDF Untuk Pengelompokan Topik Skripsi Pada Aplikasi Repository STMIK Budi Darma,” vol. 2, no. 7, pp. 414–432, 2021, [Online]. Available: <https://ejournal.seminar-id.com/index.php/tin>
- [10] S. M. Salsabila, A. Alim Murtopo, and N. Fadhilah, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [11] F. R. Irawan, A. Jazuli, and T. Khotimah, “Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors Sentiment Analysis of Gojek Users Using K-Nearest Neighbor,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.