

Analisis Sentimen Komentar Pada Saluran Youtube *Food Vlogger* Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes

Sentiment Analysis Comment on Indonesian Youtube Channel About Food Vlogger Using Naïve Bayes Algorithm

Mampe Parulian Munthe¹, Anton Siswo Raharjo Ansori², Reza Rendian Septiawan³
^{1,2,3} Universitas Telkom, Bandung
munthemampe@telkomuniversity.ac.id¹,
raharjo@telkomuniversity.ac.id², zaseptiawan@telkomuniversity.ac.id³

Abstrak

Saat ini YouTube merupakan salah satu media sosial yang paling populer. YouTube merupakan media sosial yang dapat digunakan untuk mengirim, melihat dan membagikan video. Pengguna YouTube yang menonton video YouTube dapat menyampaikan opininya melalui kolom komentar pada YouTube. Komentar yang disampaikan dapat digunakan sebagai analisis pada video YouTube tersebut. Dari analisis ini dapat dijadikan sebagai tolak ukur terhadap video yang dibuat untuk mendapatkan *feedback* dari penonton, positif atau negatif. Untuk mengatasi permasalahan klasifikasi komentar pengguna YouTube maka dirancang sebuah sistem analisis komentar berdasarkan filter YouTube dengan algoritma *Naïve Bayes*. Sistem analisis komentar pada YouTube yang dibuat akan menghasilkan klasifikasi dari komentar-komentar pengguna YouTube dengan kategori positif dan negatif. Sistem ini diharapkan dapat menjadi bahan evaluasi para pembuat konten untuk meningkatkan kualitas dari saluran YouTube-nya. Dari model yang telah dibuat dengan data komentar sebanyak 1702 komentar yang terdiri dari 848 data positif dan 854 data negatif, didapatkan hasil akurasi 90.60% dengan perbandingan data testing dan training 30:70.

Kata Kunci: Analisis Sentimen, Klasifikasi Teks, Naïve Bayes

Abstract

Nowadays YouTube is one of the most popular social media. Youtube is a social media that can be used to send, view and share videos. YouTube users who watch YouTube videos can share their opinions through the comments field on YouTube. Comments submitted can be used as an analysis on the YouTube video. From this analysis can be used as a benchmark against videos made to get feedback from viewers, positive or negative. To solve the problem of classification of comments of YouTube users, a comment analysis system based on YouTube filters was designed with a naïve bayes algorithm. YouTube's comment analysis system will result in classification of YouTube user comments in both positive and negative categories. This system is expected to be an evaluation tool for content creators to improve the quality of their YouTube channel. From the model that has been created using a dataset consists of 1702 comments with 848 positive data and 854 negative data, we obtain an accuracy of 90.60% with the ratio of testing data and training data is 30:70.

Keywords: *Sentiment Analysis, Text Classification, Naïve Bayes*

1. Pendahuluan

1.1 Latar Belakang

Saat ini penggunaan media sosial semakin meningkat. Media sosial seperti Facebook, Twitter dan YouTube menyediakan berbagai informasi untuk para penggunanya, baik berupa video maupun teks. Dari ketiga media sosial tersebut YouTube menjadi media berbagi informasi video yang terbesar. YouTube sendiri merupakan media sosial yang memungkinkan para penggunanya untuk dapat melihat, mengirim, dan berbagi video. Di Indonesia, YouTube sudah berkembang pesat. Banyak saluran YouTube dengan beragam konten video yang diunggah para pengguna YouTube. YouTube memiliki indikator penilaian dengan *rating like* atau *dislike*. Para pengguna YouTube yang menonton video pada salah satu saluran YouTube dapat menyampaikan pendapat terhadap video melalui kolom komentar jika pemilik video tidak mematikan fitur komentar. Komentar dari pengguna dapat digunakan sebagai bahan penilaian dan evaluasi terhadap kualitas dari saluran YouTube tersebut.

Untuk menghasilkan informasi dari data komentar yang ada, maka diperlukan pengolahan data dengan analisis komentar pada video YouTube. Pada penelitian ini dilakukan analisis komentar yang disampaikan oleh pengguna YouTube. Analisis ini dapat dijadikan sebagai tolak ukur terhadap video yang dibuat untuk mengetahui *feedback* yang didapat dari penonton positif atau negatif. Maka dirancanglah sebuah sistem analisis komentar pada

YouTube dengan menggunakan algoritma Naïve Bayes. Sistem analisis komentar pada YouTube yang dibuat akan menghasilkan klasifikasi dari komentar-komentar pengguna YouTube dengan kategori positif dan negatif. Diharapkan sistem ini dapat menjadi bahan evaluasi para pemilik video untuk meningkatkan kualitas dari saluran YouTube-nya.

1.2 Rumusan Masalah

1. Bagaimana cara untuk memberikan ulasan sentimen komentar pengguna YouTube terhadap video yang diunggah oleh pemilik YouTube?
2. Bagaimana performansi metode klasifikasi Naive Bayes pada sistem analisis komentar pada YouTube?

1.3 Tujuan

1. Merancang sistem analisis komentar pada YouTube dengan algoritma Naïve Bayes.
2. Mengetahui performansi sistem analisis komentar yang dirancang dengan metode klasifikasi Naïve Bayes.
3. Melakukan pengujian terhadap video YouTube dari saluran berbahasa Indonesia yang memiliki tema food vlogger.

2. Dasar Teori

2.1 Text Mining

Text mining atau analitik teks merupakan sebuah teknologi kecerdasan buatan (*artificial intelligence*, AI) yang menggunakan bahasa alami (*natural language processing*, NLP) untuk mengubah teks bebas yang tidak terstruktur dalam sebuah dokumen menjadi terstruktur yang dinormalisasikan untuk keperluan analisis pembelajaran mesin (*machine learning*, ML) [2].

2.2 Preprocessing

Preprocessing merupakan proses menghilangkan dataset yang dapat mempengaruhi tingkat akurasi. Teks perlu dibersihkan dan dikodekan menjadi nilai numerik sebelum diberikan ke model pembelajaran mesin. Proses pembersihan dan pengkodean ini disebut sebagai pemrosesan awal teks.

2.3 Labeling Kamus

Labeling Kamus merupakan proses pemberian label positif, negatif atau netral pada komentar. Kamus yang digunakan bersumber dari InSet (Indonesia Sentiment Lexicon) dan kamus yang digunakan sudah mengalami beberapa perubahan seperti penambahan atau pengurangan kata pada kamus serta penyesuaian bobot pada setiap kata [4]. Proses *Labeling Kamus* ini dilakukan secara otomatis berdasarkan kamus yang sudah diberikan bobot dengan rentang -5 sampai 5.

2.4 Analisis Sentimen

Sentimen analisis merupakan proses mengolah atau mengekstrak data secara tekstual dengan otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat tersebut [5]. Analisis sentimen ini memiliki tujuan untuk memperoleh opini dari pengguna yang terdapat pada media sosial ataupun media lainnya yang ada di internet. Analisis sentimen akan mengklasifikasikan sentimen dokumen tekstual yang diterapkan pada komentar, produk tertentu, atau topik dengan kategori positif atau negatif.

2.5 Term Weighting

Term Weighting atau pembobotan kata dilakukan setelah *preprocessing* data komentar akan dilakukan pembobotan kata untuk menghitung bobot dari setiap kata yang ada pada dokumen. Metode yang dapat digunakan salah satunya adalah TF-IDF (*Term Frequency - Inverse Document Frequency*). Bobot dari setiap data merupakan hasil perkalian dari TF dan IDF. TF merupakan frekuensi jumlah term pada data, dan IDF adalah fungsi untuk mengurangi bobot term jika kemunculannya banyak tersebar dalam dokumen [9].

$$W_{t,d} = tf_{(t,d)} \times idf_t \quad (1)$$

dengan:

$$idf_t = \log \left(\frac{N}{df_t} \right) + 1 \quad (2)$$

$$tf(t,d) = \frac{f_{t,d}}{\text{number of word in } d} \quad (3)$$

dengan:

- $f_{t,d}$ =Frekuensi kemunculan term
- N =Jumlah kata pada dokumen
- idf_t = Inverse Document Frequency
- dft =Total term pada dokumen

2.6 Naïve Bayes Classifier

Naïve Bayes. Naïve Bayes merupakan metode klasifikasi yang menerapkan teorema Bayes yang memungkinkan untuk menghitung ketidakpastian tentang model dengan prinsip menentukan probabilitas[10]. Secara umum Naïve Bayes memiliki teorema sebagai berikut [5] :

$$P(C_j|W_i) = \frac{P(W_i|C_j)P(C_j)}{P(W_i)} \quad (4)$$

dengan:

- $P(C_j|W_i)$: Peluang kategori j , ketika terdapat kemunculan kata i .
- $P(W_i|C_j)$: Peluang kata i masuk kedalam kategori.
- $P(C_j)$: Peluang kemunculan kategori j .
- $P(W_i)$: Peluang kemunculan kata.
- Untuk menghitung $P(C_j|W_i)$, dapat dilakukan dengan rumus berikut:

$$P(W_i|C_j) = \frac{N_{ic}+1}{N_c+V} \quad (5)$$

dengan:

- N_{ic} : Jumlah dokumen latih dengan fitur atribut W_i dengan kategori C_j
- N_c : Jumlah dokumen dari kategori C_j
- V : Jumlah kategori

2.7 Evaluasi Performansi

Evaluasi performansi dilakukan dengan melihat nilai *Confusion Matrix*. Evaluasi performansi ini perlu dilakukan untuk mengukur keakuratan metode dan kesesuaian dari sistem yang dibuat. Evaluasi performansi ini dilakukan dengan uji performansi *accuracy*, *precision*, *recall*, *f-measure*[10]. Perhitungan performansi menggunakan confusion matrix seperti pada tabel dibawah:

Tabel 2.1 *Confusion matrix* untuk evaluasi performansi.

		Nilai Aktual	
		Negatif	Positif
Nilai Prediksi	Negatif	True Negative (TN)	False Positif (FP)
	Positif	False Negatif (FN)	True Positif (TP)

Ada beberapa kondisi pada tabel diatas yaitu: True Positive (TP) saat nilai prediksi positif dan nilai sebenarnya positif, False Positive (FP) saat nilai prediksi positif dan nilai sebenarnya negative, False Negative (FN) saat nilai prediksi negatif dan nilai sebenarnya positif, True Negative (TN) saat nilai prediksi negatif dan nilai sebenarnya negatif. Berikut merupakan persamaan dari *accuracy*, *precision*, *recall*, dan *f-measure*:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

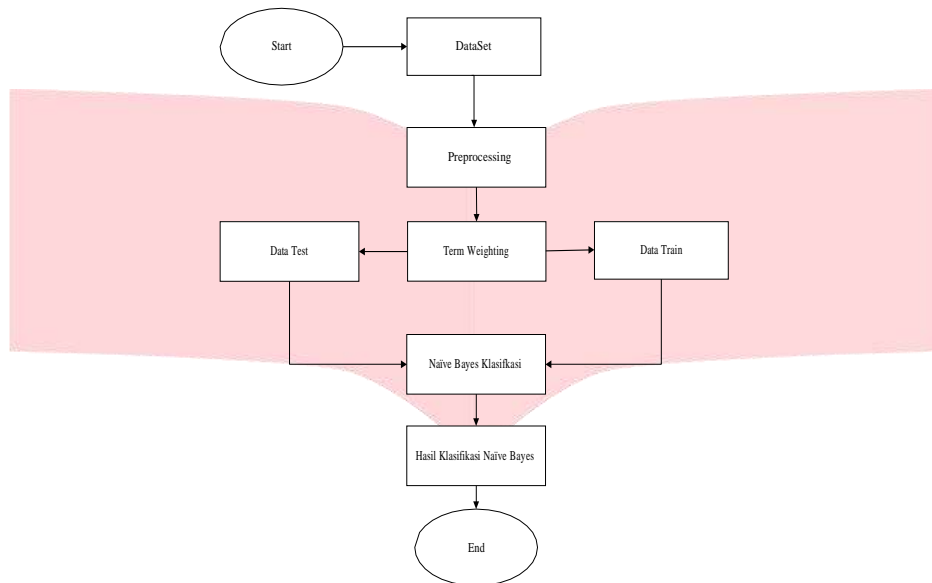
$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (9)$$

3. Perancangan

3.1. Sistem yang Dibangun

Dalam penelitian ini sistem yang akan dibangun dapat mengklasifikasikan komentar pada YouTube. Gambaran umum dari sistem yang dibangun pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 3.1 Gambaran umum sistem.

3.2 Dataset

Dataset yang digunakan pada penelitian ini diperoleh dengan cara melakukan *scraping* pada komentar YouTube. Dari hasil data yang dikumpulkan, untuk pengujian data training dan test akan digunakan sebanyak 1702 komentar, dengan jumlah data positif sebanyak 848 data dan negatif sebanyak 854 data. Pemberian label pada data komentar dilakukan dengan otomatis berdasarkan kamus yang telah dibuat.

3.3 Preprocessing

Preprocessing data untuk menghilangkan *noise* pada data yang dapat mempengaruhi hasil akhir data tersebut. Tahapan *preprocessing* ini dilakukan sebelum melakukan klasifikasi data. Tahapan *preprocessing* ini meliputi: Case Folding, Tokenizing, Normalisasi Kata, Filtering dan Stemming.

3.4 Labeling Komentar

Labeling Komentar akan diberikan setelah data komentar YouTube melalui tahapan *preprocessing*. Selanjutnya akan diberikan bobot pada setiap kata sesuai dengan kamus yang telah dibuat dengan rentang -5 sampai 5. Kamus yang telah dibuat terdiri dari 3408 kata positif dan 5190 kata negatif.

3.5 Term Weighting (TF-IDF)

Pembobotan kata merupakan proses pemberian nilai pada setiap term yang ada pada data komentar YouTube setelah dilakukan *preprocessing* dan labeling. Pemberian bobot terhadap term dilakukan dengan metode Term Frequency – Inverse Document Frequency (TF-IDF). Contoh perhitungan TF-IDF pada sebuah dokumen, dapat dilihat pada contoh berikut ini: Dokumen 1 = “makan enak saya suka”, Dokumen 2 = “makan ayam pagi hari”.

Tabel 3.1 Perhitungan TF-IDF.

dokumen	term		tf		idf		tf * idf	
	d1	d2	d1	d2	d1	d2	d1	d2
Makan	1	1	0.25	0.25	1	1	0.25	0.25
Enak	1	0	0.25	0	1.693	0	0.423	0
Saya	1	0	0.25	0	1.693	0	0.423	0
Suka	1	0	0.25	0	1.693	0	0.423	0
Ayam	0	1	0	0.25	0	1.693	0	0.423
Pagi	0	1	0	0.25	0	1.693	0	0.423
Hari	0	1	0	0.25	0	1.693	0	0.423

3.6 Klasifikasi Naïve Bayes

Data komentar YouTube yang sudah melalui proses *preprocessing*, *labeling* dan *term weighting* (TF-IDF) akan menjadi input pada proses *test* dan *training*. Selanjutnya data akan melalui tahap pengklasifikasian yang menerapkan algoritma *Multinomial Naïve Bayes*. Pada proses pengklasifikasian dengan Naïve Bayes akan dihitung nilai *probability* dari *class* dari dokumen yang diberikan, nilai *probability* dari *class* didapat dari *priors* yang dicari dibagi dengan jumlah total *priors* yang ada. Kemudian akan dicari nilai dari *conditional probability*. Nilai dari *conditional probability* dicari dengan menghitung kata jumlah kata yang muncul pada *class* dibagi dengan jumlah semua kata pada *class*, lalu ditambahkan satu untuk *smoothing*. Lalu untuk mencari *class* dari data *testing*, setelah didapat nilai *priors* dan *conditional probability* dilakukan dengan menghitung nilai *priors* dikali dengan nilai *conditional probability*. Untuk menentukan *class* data *testing* akan dibanding nilai yang didapat, nilai yang lebih besar akan menjadi *class* data *testing* tersebut. Berikut merupakan contoh perhitungan klasifikasi menggunakan Naïve Bayes Classifier:

Tabel 3. 2 Perhitungan Naïve Bayes.

No	Dokumen	Label
1	makan enak saya suka	Positif
2	makan ayam pagi hari	Netral
3	makan mahal	Negatif
4	ayam enak saya suka	?

Pada tabel 3.2 di atas merupakan contoh dari data yang telah diberikan label, dokumen ke-1 positif, dokumen ke-2 netral dan dokumen ke-3 negatif, pada dokumen ke-4 akan diberikan label. Untuk memberikan label pada dokumen ke-4 akan digunakan metode klasifikasi dengan Naïve Bayes. Untuk tahapan klasifikasi adalah sebagai berikut:

- Tahap pertama dilakukan dengan memberikan pemobotan kata pada setiap dokumen dengan menggunakan *Term Frequency*
- Menghitung probabilitas *prior* setiap kelas, yang menjadi kelas ada 3 yaitu: positif, negatif dan netral.

$$-P(\text{positif}) = \frac{1}{3} = 0.33 \quad -P(\text{negatif}) = \frac{1}{3} = 0.33 \quad -P(\text{netral}) = \frac{1}{3} = 0.33$$
- Menghitung *conditional probabilities* setiap *term* dari semua dokumen.

- Probabilitas kata "makan" $P(\text{makan} \text{positif}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{makan} \text{netral}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{makan} \text{negatif}) = \frac{1+1}{2+8} = \frac{2}{10} = 0.16$ - Probabilitas kata "saya" $P(\text{makan} \text{positif}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{makan} \text{netral}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.08$ $P(\text{makan} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.2$ - Probabilitas kata "ayam" $P(\text{makan} \text{positif}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.08$ $P(\text{makan} \text{netral}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{makan} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.1$ - Probabilitas kata "hari" $P(\text{makan} \text{positif}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{makan} \text{netral}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.08$ $P(\text{makan} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.1$	- Probabilitas kata "enak" $P(\text{enak} \text{positif}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{enak} \text{netral}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.08$ $P(\text{enak} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.1$ - Probabilitas kata "suka" $P(\text{enak} \text{positif}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{enak} \text{netral}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.08$ $P(\text{enak} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.1$ - Probabilitas kata "pagi" $P(\text{enak} \text{positif}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.08$ $P(\text{enak} \text{netral}) = \frac{1+1}{4+8} = \frac{2}{12} = 0.16$ $P(\text{enak} \text{negatif}) = \frac{0+1}{2+8} = \frac{1}{10} = 0.1$ - Probabilitas kata "mahal" $P(\text{enak} \text{positif}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.16$ $P(\text{enak} \text{netral}) = \frac{0+1}{4+8} = \frac{1}{12} = 0.16$ $P(\text{enak} \text{negatif}) = \frac{1+1}{2+8} = \frac{2}{10} = 0.2$
--	---
- Setelah didapat nilai *prior* dan nilai *conditional probabilities* maka untuk menentukan kelas pada dokumen ke-4 "ayam enak saya suka" dapat dilakukan dengan mengalikan semua nilai peluang.

$-P(d4 \text{positif})$ $= 0.33 \times 0.08 \times 0.16 \times 0.16 \times 0.16$ $= 0.00067584$	$-P(d4 \text{netral})$ $= 0.33 \times 0.16 \times 0.08 \times 0.08 \times 0.08$ $= 0.0000270336$	$-P(d4 \text{negative})$ $= 0.33 \times 0.1 \times 0.1 \times 0.2 \times 0.1$ $= 0.000066$
---	--	--

Setelah didapatkan nilai peluang dari setiap kelas, lalu akan dibandingkan nilai $P(d4 | positif)$, $P(d4 | netral)$ dan $P(d4 | negatif)$. Dari hasil yang didapatkan nilai 0.00067584 yang lebih besar maka pada dokumen ke-4 “ayam enak saya suka” memiliki kelas positif.

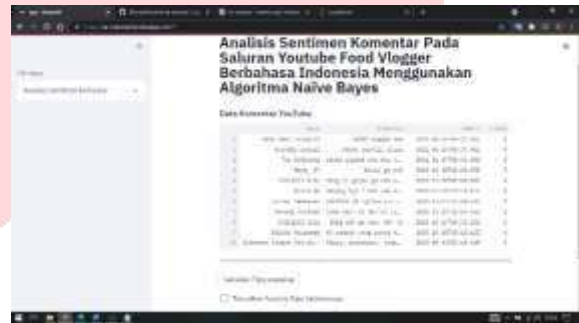
4. Pengujian dan Implementasi

4.1 Implementasi aplikasi web

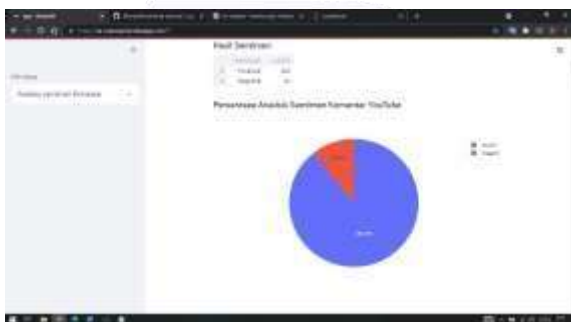
Desain perangkat lunak yang akan dibangun pada penelitian tugas akhir ini menggunakan aplikasi web yang dibangun dengan menggunakan bahasa pemrograman Python dan framework Streamlit untuk memproses data dan menampilkan data hasil dari klasifikasi.



Gambar 4. 1 Halaman awal aplikasi web.



Gambar 4. 2 Halaman analisis sentimen.



Gambar 4. 3 Halaman analisis sentimen.



Gambar 4. 4 Halaman tentang aplikasi web.

4.2 Pengujian Pelatihan Model dan Akurasi Sistem

Tabel 4. 1 Pengujian model dan akurasi sistem.

Test-Size	Train-Size	N-Gram	Alpha	Akurasi
20%	80%	(1.2)	46	88.26
30%	80%	(1.2)	46	90.60
40%	60%	(1.2)	46	88.54
50%	50%	(1.2)	46	88.24
60%	40%	(1.2)	46	88.16
70%	30%	(1.2)	46	87.33
80%	20%	(1.2)	46	86.27
20%	80%	(1.3)	41	87.97
30%	70%	(1.3)	41	90.60
40%	60%	(1.3)	41	88.54
50%	50%	(1.3)	41	88.01
60%	40%	(1.3)	41	86.25
70%	30%	(1.3)	41	86.99
80%	20%	(1.3)	41	86.04

Pada tabel 4.1 melakukan pelatihan model dilakukan beberapa kondisi dengan mengubah parameter pada model agar didapatkan nilai akurasi yang tinggi. Dalam hal ini beberapa pengaturan model yang dapat dilakukan yaitu dengan mengubah n-gram pada model, mengubah porsi dari data test dan training, memberikan nilai optimal pada nilai alpha Naive Bayes. Pada pengujian ini data yang digunakan adalah kumpulan data yang telah di-scraping dan telah dilakukan preprocessing pada tahapan sebelumnya dan diberikan labeling secara otomatis berdasarkan kamus. Berdasarkan pengujian model didapatkan skor akurasi tertinggi sebesar 90.60 %, saat menggunakan kombinasi parameter *test-size* 30%, *train-size* 70%, *alpha* 46, dan *n-gram* 1.2.

4.3 Pengujian Akurasi Dengan Confusion Matrix

Mengukur kinerja suatu model merupakan suatu hal penting dalam sebuah *machine learning* sehingga dapat menjadi sebuah pertimbangan apakah model itu baik atau tidak. Untuk mengukur kinerja model salah satu teknik yang dapat dilakukan yaitu dengan *Confusion Matrix*.

Tabel 4. 2 Pengujian Akurasi Confusion Matrix.

		Nilai Aktual	
		Negatif	Positif
Nilai Prediksi	Negatif	233	25
	Positif	23	230

Setelah didapatkan nilai *Confusion Matrix*, maka nilai dari *accuracy*, *precision*, *recall*, dan *f-measure* dapat dihitung.

- Menghitung $accuracy = \frac{230+233}{230+25+233+23} = 0.9060$
- Menghitung $precision = \frac{230}{230+25} = 0.9019$
- Menghitung $recall = \frac{230}{230+23} = 0.9090$.
- Menghitung $F-Measure = \frac{2 \cdot (0.9019 \cdot 0.9090)}{0.9019 + 0.9090} = 0.9055$

4.4 Pengujian Analisis Sentimen Pada Video

Pengujian analisis sentimen dilakukan setelah *training* model dan disimpan dengan *library pickle*. Selanjutnya model yang telah disimpan akan diimplementasikan pada beberapa video yang menjadi pengujian pada penelitian ini. Video yang akan diuji bukan dari pengujian data *training* dan *test* pada model sebelumnya. Pada pengujian ini akan dibandingkan hasil prediksi dari model yang dibangun dengan perhitungan secara manual.

Tabel 4. 3 Pengujian analisis sentimen video YouTube I.

URL Video	https://www.youtube.com/watch?v=QsRK2gEuJDw	
Judul Video	MANTAN CHEF HOTEL BINTANG 5 JUALAN SUSHI DI DALEM GANG !!	
Jumlah Komentar	339	
	Hasil Model	Hasil Manual
Positif	177	187
Negatif	68	58

URL Video	https://www.youtube.com/watch?v=wZYSI8jP59g	
Judul Video	TUKANG MARAH VIRAL JUALAN BAKMI PENTOL, MAU BELI MALAH DI OMELIN !!	
Jumlah Komentar	232	
	Hasil Model	Hasil Manual
Positif	97	110
Negatif	62	49

5. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan pada Analisis Sentimen Komentar Pada Saluran Youtube Food Vlogger Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes maka diperoleh kesimpulan yaitu:

1. Sistem analisis komentar pada YouTube dengan algoritma Naïve Bayes yang dirancang dapat mengklasifikasikan komentar ke dalam kategori positif atau negatif.
2. Klasifikasi sentimen dengan menggunakan dataset yang dipilih secara acak dengan topik food vlogger dengan jumlah data sebanyak 1702 komentar, dengan jumlah positif 848 data dan negatif 854 data didapatkan hasil akurasi 90.06% dengan pembagian data *testing* dan data *training* sebanyak 30:70.
3. Model yang telah dibuat diujikan pada saluran YouTube milik Ken & Grat dengan topik *food vlogger*. Dari dua video yang diuji untuk video pertama memiliki 72.2% positif dan 27.8% negatif, untuk video kedua memiliki 61% positif dan 39% negatif perhitungan ini merupakan dengan model, sedangkan untuk perhitungan secara manual didapat 76.3% positif dan 23.7% negatif, untuk video kedua 69.2 % positif dan 30.8% negatif.

REFERENSI

- [1] D. Novianty, "YouTube Rajai Media Sosial di Indonesia." <https://www.suara.com/tekno/2021/02/15/153000/youtube-rajai-media-sosial-di-indonesia> (accessed Nov. 05, 2020).
- [2] Linguamatics, "What is Text Mining, Text Analytics and Natural Language Processing? Linguamatics." <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing> (accessed Nov. 27, 2020).
- [3] M. Ibad, "ANALISIS SENTIMEN POSITIF DAN NEGATIF KOMENTAR VIDEO YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES – SUPPORT VECTOR MACHINE (NBSVM) CLASSIFIER," p. 2, 2016.
- [4] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2018, doi: 10.1109/IALP.2017.8300625.
- [5] P. Y. Saputra, D. H. Subhi, and F. Z. A. Winatama, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes," *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, 2019, doi: 10.33795/jip.v5i4.259.
- [6] "YouTube Data API | Google Developers." <https://developers.google.com/youtube/v3> (accessed Oct. 15, 2020).
- [7] E. Hariyadi, "Membuat Proyek Pertama di Heroku - Codepolitan." <https://www.codepolitan.com/membuat-proyek-pertama-heroku-58b872c6217eb> (accessed Oct. 15, 2020).
- [8] Stackabuse, "Introduction to Web Scraping with Python." <https://stackabuse.com/introduction-to-web-scraping-with-python/> (accessed Nov. 29, 2020).
- [9] W. Scott, "TF-IDF from scratch in python on real world dataset. | by William Scott | Towards Data Science." <https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089> (accessed Nov. 20, 2020).
- [10] D. Rizkia, H. Tiasari, I. Asror, Y. R. Murti, and M. Kom, "Analisis Sentimen Mengenai Performansi Saluran YouTube dengan Metode Naïve Bayes."