

Analisis Sentimen Terhadap Dampak Inflasi di Indonesia Menggunakan Metode Multinomial Naïve Bayes

Sentiment Analysis of the Impact of Inflation in Indonesia Using the Naïve Bayes Multinomial Method

Aldi Tri Wijaya¹, Arief Hermawan²

¹Informatika, Universitas Teknologi Yogyakarta

E-mail: ¹aldi.5200411242@student.uty.ac.id, ²ariefdb@uty.ac.id

Abstrak

Inflasi yang terjadi di seluruh dunia termasuk negara Indonesia mengakibatkan masalah bagi masyarakat maupun negara, terjadinya ketidakstabilan ekonomi negara dan meningkatnya pengangguran dikarenakan banyaknya pengurangan karyawan kerja, selain itu dengan kelangkaan minyak terutama bahan bakar kendaraan salah satu faktor yang mendorong kenaikan inflasi sehingga mengakibatkan harga-harga barang dan sandang mengalami kenaikan yang signifikan, dari persoalan tersebut menimbulkan opini-opini masyarakat pada jejaring sosial khususnya *twitter* dari timbulnya masalah inflasi. Maka dibuatkan sistem untuk melakukan analisis sentimen dari masalah kondisi Inflasi Indonesia. Proses sistem memerlukan data *tweet* dengan jumlah 1725 tweet hasil dari proses *crawling* data *twitter* menggunakan bantuan *library python* yaitu *tweepy* pada masing-masing kata kunci inflasi. Algoritma *naïve bayes clasiffier* dengan model *multinomialnb* sebagai metode untuk melakukan klasifikasi. Hasil klasifikasi dari *search key* inflasi indonesia mengandung beberapa kelas yaitu positif, negatif dan netral. Proses penelitian diuji menggunakan beberapa skenario pembagian data yaitu 90:10, 80:20, 70:30, untuk tingkat akurasi terbaik dihasilkan menggunakan skenario 90% data train dan 10% data test mendapatkan hasil akurasi 75,5%, *precision* 75%, *f1-score* 75% dan *recall* 74%.

Kata kunci: Inflasi, Klasifikasi, *Naïve Bayes Classifier*, *MultinomialNb*, Sentimen.

Abstract

Inflation that occurs throughout the world, including Indonesia, causes problems for the community and the state, the occurrence of state economic instability and increased unemployment due to the large reduction in work employees, in addition to the scarcity of oil, especially vehicle fuel, one of the factors that drives the increase in inflation resulting in the prices of goods and clothing experiencing a significant increase, from this problem raises opinions People on social networks, especially Twitter, from the emergence of inflation problems. So a system was created to conduct sentiment analysis of the problem of Indonesia's inflation conditions. The system process requires tweet data with a total of 1725 tweets resulting from the process of crawling twitter data using the help of a python library, namely tweepy on each inflation keyword. Naïve bayes clasiffier algorithm with multinomialnb model as a method for classifying. The classification results of the Indonesian inflation search key contain several classes, namely positive, negative and neutral. The test process was carried out with several data sharing scenarios, namely 90:10, 80:20, 70:30, for the best level of accuracy produced using the 90% data train scenario and 10% data test getting 75.5% accuracy results, 75% precision, 75% f1-score and 74% recall

Keywords: *Inflation, Classification, Naïve Bayes Classifier MultinomialNb, Sentiment.*

1. PENDAHULUAN

Inflasi adalah proses kenaikan umum sebagian besar komoditas barang selama periode tertentu, mempengaruhi kondisi individu masyarakat, pebisnis dan pemerintah [1]. Kenaikan harga hanya beberapa barang bukan disebut inflasi [2], kecuali jika kenaikan sama rata. Inflasi dapat disebut masalah ekonomi yang tidak dapat dilalaikan begitu saja, maka dapat menyebabkan banyak masalah baik dari negara Indonesia maupun masyarakat. Inflasi yang cukup tinggi, sangat penting diawasi oleh pemerintah karena akan berdampak pada sektor ekonomi yang bisa membuat ketidakstabilan ekonomi dan mengganggu *stabilitas* sosial dan politik [3] dan dikhawatirkan akan menambah jumlah garis kemiskinan masyarakat [4]. Faktor-faktor yang dapat mempengaruhi kenaikan inflasi adalah harga bahan bakar minyak dunia dan ada gangguan panen atau banjir [5], kemudian fenomena perang antara Rusia dan Ukraina berdampak pada harga minyak mentah [6] sehingga perekonomian dunia mengalami dampak dari perang termasuk Indonesia dimana pihak Rusia membatasi penjualan minyak terhadap beberapa negara dunia, sehingga menimbulkan kekacauan perekonomian dunia yang mengakibatkan kenaikan inflasi di berbagai negara, dengan faktor-faktor tersebut pemerintah diharapkan mampu membuat sebuah kebijakan yang mampu menekan angka inflasi yang rendah bagi Indonesia, jika inflasi rendah akan mendorong ekonomi [7] dan mengurangi permasalahan negatif bagi negara, karena tingkat inflasi mencerminkan kestabilan perekonomian negara, yang akan membuat masyarakat di Indonesia menjadi sengsara atau lebih baik, Indonesia sendiri tingkat inflasi tertinggi selama periode 2000-2021, dengan tertinggi pada tahun 2006 yaitu 13,11 % [8] yang disebabkan oleh kenaikan harga BBM dan kenaikan suku bunga riil. Dari persoalan-persoalan terhadap kondisi inflasi di Indonesia menimbulkan beberapa opini-opini masyarakat berkaitan dengan inflasi tersebut khususnya di media sosial twitter.

Penentuan sentimen opini atau umpan balik dapat dilakukan dan dipahami secara manual tetapi jika data yang digunakan sangat banyak maka membutuhkan waktu yang lama [9] untuk menentukan sentimen tersebut, maka dari itu dari sistem analisis sentimen dapat mempermudah pengguna dalam mengklasifikasi setiap kelas sentimen positif, negatif maupun netral [10] secara otomatis, sehingga pengguna dapat mengetahui opini tentang topik inflasi di Indonesia secara rinci, analisis sentimen ini akan memberikan informasi berharga bagi pemerintah, pengambilan kebijakan dan pelaku pasar untuk merespon dan mengantisipasi perubahan harga dan dampak inflasi secara lebih efektif.

Metode klasifikasi *naïve bayes* sangat populer untuk mengklasifikasi sebuah teks menggunakan *probabilitas* [11], metode klasifikasi *multinomial naïve bayes* akan digunakan pada penelitian ini sebagai metode untuk mengetahui performa saat melakukan sentimen analisis secara otomatis oleh sistem dan metode *naïve bayes* dengan model *multinomial* belum pernah digunakan untuk kasus data *tweet* dengan topik inflasi. Penelitian yang mirip mengandung kata kunci inflasi pernah dilakukan oleh Puspitasari, dkk yaitu analisis sentimen terhadap inflasi pasca covid 19 berdasarkan *twitter*. Penelitiannya menggunakan algoritma *K-NN* dan *SVM*, menghasilkan akurasi tertinggi dengan nilai akurasi 79% pada algoritma *SVM*, untuk model *K-NN* menggunakan skenario pembagian data 90:10 mendapatkan akurasi 54% [12], sehingga penelitian ini akan menggunakan algoritma klasifikasi *multinomial naïve bayes* sebagai pembandingan dari penelitian sebelumnya.

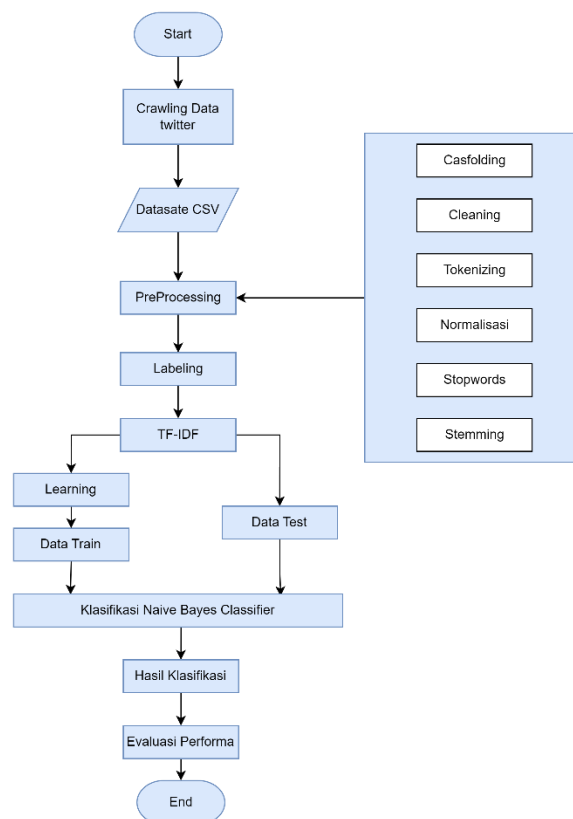
Analisis sentimen menggunakan *naïve bayes* berdasarkan data *twitter* juga sempat dilakukan untuk menganalisis sentimen terhadap kenaikan iuran BPJS Kesehatan pada *twitter* menggunakan algoritma *naïve bayes classifier* menghasilkan akurasi sebesar 90,96 % [13], implementasi analisis sentimen masyarakat mengenai kenaikan harga BBM dengan metode *gaussian naïve bayes* menghasilkan akurasi terbesar yaitu 68% [14], analisis sentimen pengguna transportasi online terhadap layanan grab Indonesia menggunakan *multinomial naïve bayes classifier* menghasilkan akurasi akhir 86,57% [15], sentimen analisis belajar online di twitter menggunakan *naïve bayes* menghasilkan akurasi sebesar 74,08% [16], sistem analisis sentimen pada ulasan produk menggunakan metode *naïve bayes* menghasilkan akurasi tertinggi yaitu 77,78% [17], analisis sentimen pada *twitter* terhadap mobil listrik menggunakan algoritma *naïve bayes* berhasil

mendapatkan akurasi 87,43% [18].

Berdasarkan persoalan dan penelitian sebelumnya belum pernah dilakukan analisis sentimen terhadap topik inflasi dengan algoritma *multinomial naïve bayes*, analisis sentimen terhadap inflasi memiliki peran yang sangat penting dalam pengambilan keputusan efektif dan pengembangan kebijakan yang tepat, maka penelitian ini bertujuan memberikan pemahaman tentang sentimen masyarakat terhadap dampak inflasi di Indonesia menggunakan *multinomial naïve bayes* dan *splitting data* sebagai acuan perbandingan akurasi dari sistem ini, sehingga penelitian ini diharapkan dapat terwujud langkah-langkah kebijakan yang efektif dalam mengelola inflasi dan memperkuat stabilitas negara.

2. METODE PENELITIAN

Perancangan sistem disusun dengan metode penelitian untuk membuat sistem analisis sentimen sesuai tahap-tahapan yang ada pada proses membangun sistem, berikut alur penelitian pada Gambar 1.



Gambar 1 Proses Penelitian

2.1 Crawling Twitter

Proses *crawling* yaitu pengambilan data berupa *tweet* dari jejaring sosial *twitter*. Tipe pengumpulan data menggunakan tipe *cursor*, yaitu mengambil data dengan beberapa kriteria seperti berdasarkan *time range tweet* dan *retweet* pengambilan data dilakukan dalam rentang bulan september sampai oktober 2022, proses *crawling* data ada beberapa cara untuk melakukan proses pengambilan tweet pada twitter API seperti menggunakan *tweepy*, *TWINT* dan *snsrape*[19], penulis menggunakan *library* pada *python* yaitu *tweepy* untuk proses pengambilan data tweet di twitter, pada proses ini menghasilkan kumpulan-kumpulan data tweet yang disimpan berupa dataset *csv*.

2.2 PreProcessing

Preprocessing dilakukan untuk menghilangkan karakter-karakter ataupun data teks yang yang tidak sesuai sehingga dapat membantu pemrosesan data menjadi akurat, tujuan dari *preprocessing text* agar data yang didapat terstruktur dan mudah dalam melakukan proses pengolahan data [20] pada proses *preprocessing* terdapat beberapa tahap pada penelitian ini yaitu:

1. *Casefolding* yaitu mengubah teks di dalam dokumen menjadi huruf *lowercase* semua.
2. *Cleaning* yaitu menghilangkan karakter khusus seperti *symbol* dan emoticon yang nantinya tidak dibutuhkan dan tidak berpengaruh terhadap bobot sentimen.
3. *Tokenizing* yaitu proses memecah sebuah kumpulan kalimat menjadi kumpulan perkata.
4. Normalisasi yaitu menormalkan kata-kata sesuai dengan kamus kata slang, kata singkatan, dengan kamus yang digunakan yaitu dataset csv untuk menormalkan kata tersebut.
5. *Stopwords* yaitu menghapus kata kata sambung seperti kata di, yang, ke, dll yang tidak berpengaruh pada proses sentimen analisis.
6. *Stemming* yaitu mengubah kata yang berimbuhan menjadi ke kata dasar atau teknik yang digunakan untuk mendapatkan sebuah kata dasar.

2.3 Pembobotan tf-idf

Pembobotan *tf-idf* yaitu proses menghitung banyaknya kata dan memberikan bobot pada setiap dokumen. Pembobotan dilakukan untuk proses klasifikasi *naïve bayes*, teknik *tf-idf* ini merupakan teknik *statistic numeric* yang memberikan nilai pada kata untuk dokumen tertentu pada *korpus*. Berikut merupakan rumus menghitung *tf-idf* [21].

$$tf_{t,d} = \frac{f_d(t)}{\max f_d(d)} \quad (1)$$

Keterangan:

$tf_{t,d}$ = term frequency

$f_d(t)$ = jumlah kemunculan kata

$\max f_d(d)$ = jumlah kata dalam dokumen

$$idf(t, D) = \log\left(\frac{N}{N_t}\right) \quad (2)$$

Keterangan:

$idf(t, D)$ = nilai *idf* pada suatu kata

N = jumlah total dokumen dalam korpus

N_t = jumlah dokumen yang mengandung kata kunci

$$tfidf_{t,d} = tf_{t,d} * idf(t, D) \quad (3)$$

Keterangan:

$tfidf_{t,d}$ = bobot term

2.4 Klasifikasi Naïve Bayes

Naïve bayes yaitu kelompok dari jenis algoritma *supervised classification* dimana dalam menentukan sebuah klasifikasi mengambil nilai probabilitas yang paling tinggi untuk hasil output klasifikasi. Penelitian ini menggunakan salah satu model yang ada di *naïve bayes* yaitu model *multinomialNB* karena model ini bekerja berdasarkan sekumpulan bukti dan kelas dengan melakukan pelatihan terhadap sebagian dataset. Perhitungan *naïve bayes* dilakukan sebagai berikut.

Posterior Probability:

$$C_{MAP} = \underset{c \in C}{\operatorname{argmax}} P(c|d) = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{i=1}^n P(w_i|c) \quad (4)$$

Prior probability:

$$P_c = \frac{N_c}{N} \quad (5)$$

Conditional probability:

$$P(w|c) = \frac{\operatorname{count}(w_i,c)+1}{\operatorname{count}(c)+|v|} \quad (6)$$

Keterangan:

- C_{MAP} : Nilai *Max A posteriori* nilai yang tertinggi probabilitas setiap sentimen.
- $P_{(c|d)}$: probabilitas kelas c yang bersyarat d
- $P_{(w|c)}$: Probabilitas kemunculan kata (Wi) di dokumen (c).
- $P_{(c)}$: Nilai probabilitas sebelumnya pada kelas c.
- $N_{(c)}$: Jumlah dokumen pada kelas (c).
- $\operatorname{count}_{(w|c)}$: Jumlah kata pada kelas (c).

2.5 Evaluasi

Proses *evaluasi* dilakukan untuk mengetahui hasil pelatihan dari klasifikasi model *naïve bayes* dengan menentukan *score accuracy, precision, recall, f1-score* dengan tujuan untuk melihat keakuratan model dalam melakukan mengklasifikasi, untuk menentukan nilai-nilai tersebut dibutuhkan sebuah *confusion matrix* karena dari matrik tersebut bisa dianalisis keakuratan dari prediksi dari sebuah model klasifikasi yang dibuat. Berikut tabel *confusion matrix* Tabel 1.

Tabel 1 *Confusion matrix*

	Prediksi positif	Prediksi netral	Prediksi negatif
Aktual positif	TP	FN	FN
Aktual negatif	FN	TN	FN
Aktual positif	FP	FP	TN

Keterangan:

- TP = data diprediksi benar “positif” oleh model.
- TN = data diprediksi benar “netral” atau “negatif” oleh model.
- FN = data diprediksi benar “netral” atau “negatif” oleh model namun sebenarnya kelas “positif”
- FP = data diprediksi benar “netral” atau “positif” oleh model namun sebenarnya kelas “netral” atau “negatif”

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset diperoleh menggunakan metode *crawling* atau *scraping*, dimana data berupa kumpulan *tweet* khusus berbahasa Indonesia yang bersumber dari media sosial *twitter*, dengan *search key* berupa kata “inflasi indonesia” menggunakan *API Twitter* yang didapatkan dengan cara mengirim permohonan kepada pihak *twitter development* dan berhasil mendapatkan jumlah data sebanyak 1725 *tweet* yang kotor yang kemudian disimpan pada file csv, pengambilan data *tweet* menggunakan *time range* pada bulan Oktober sampai September 2022. Contoh data komentar masyarakat Indonesia terhadap inflasi yang berhasil dilakukan *scraping* terdapat pada Tabel 2.

Table 2 Contoh tweet masyarakat

Waktu	Id	Username	Tweet
2022-09-01 03:02:10+00:00	158727879855 0335488	Muhamma d Gandy	"b'https://t.co/KniNSTIijq\n ALHAMDULLILAH inflasi INDONESIA turun dan terjagga
2022-10-01 03:00:08+00:00	158727828833 4254083	antaranews. com	Bbm tidak terjaga, inflasi meroket P\xe2\x80\xa6 https://t.co/fNqtsyERRz"
2022-10-29 14:00:02+00:00	158635719066 6977282	kriptonews	"ayo investasi kripto di #pluang \xe2\x80\xa6 https://t.co/JSteuCsQe2"

3.2 Pre-Processing

Proses *preprocessing* untuk membersihkan *tweet* dari *noise* yang akan berpengaruh terhadap proses saat perhitungan *TF-IDF*, dan hasil analisis sentimen. Berikut contoh hasil *preprocessing*, pada Tabel 3.

Tabel 3 Hasil *Preprocessing*

Tweet asli	"b'https://t.co/KniNSTIijq\n ALHAMDULLILAH inflasi INDONESIA turun dan terjaga
Casefolding	"b'https://t.co/kninstiijq\n alhamdullilah inflasi indonesia turun dan terjaga
Cleaning	alhamdullilah inflasi indonesia turun dan terjaga
Tokenizing	['alhamdullilah', 'inflasi', 'indonesia', 'turun', 'dan', 'terjagga']
Normalisasi	['alhamdullilah', 'inflasi', 'indonesia', 'turun', 'dan', 'terjaga']
Stopwords	['alhamdullilah', 'inflasi', 'indonesia', 'turun', 'terjaga']
Stemming	['alhamdullilah', 'inflasi', 'indonesia', 'turun', 'jaga']

Proses *preprocessing* ini akan menghasilkan sebuah kumpulan banyak kata, sehingga untuk melihat kata-kata yang sering muncul tersebut diperlukan visualisasi data dengan *worldcloud*. Berikut gambaran visualisasi *worldcloud* dari data penelitian pada Gambar 2.



Gambar 2 Visualisasi *Worldcloud*

3.3 Labeling

Proses pelabelan dilakukan secara manual dan menggunakan kelas sentimen berupa kelas positif, negatif dan netral dengan data yang digunakan berjumlah 526 (hasil *preprocessing*). Data yang termasuk dalam kelas positif termasuk data *tweet* pernyataan senang terhadap kondisi inflasi. Data negatif adalah data *tweet* yang menyatakan kurang senang terhadap kondisi inflasi. Data netral adalah data *tweet* yang tidak menunjukkan ekspresi positif dan negatif pada kondisi inflasi.

Berikut contoh hasil pemberian labeling, dapat dilihat pada Tabel 4.

Tabel 4 Labeling

Tweet	Label
alhamdulillah inflasi indonesia turun terjaga	positif
bbm tidak terjaga inflasi meroket	negatif
ayo investasi kripto peluang	netral

3.4 Pembobotan TF-IDF

Proses pembobotan digunakan untuk menghitung bobot tiap kata dalam suatu dokumen dengan menghasilkan nilai akhir berupa bobot TF-IDF yang nantinya bobot tersebut digunakan untuk menghitung probabilitas dengan naive bayes. Berikut proses perhitungan bobot TF-IDF yang dijabarkan dengan Tabel 5 berikut.

Tabel 5 Perhitungan TF, DF, dan IDF

Term	TF			DF	IDF = jumlah dokumen / DF
	d1	d2	d3		
alhamdulillah	1	0	0	1	$\text{Log}(3/1) = 0,4771$
inflasi	1	1	0	2	$\text{Log}(3/2) = 0,1760$
indonesia	1	0	0	1	$\text{Log}(3/1) = 0,4771$
turun	1	0	0	1	$\text{Log}(3/1) = 0,4771$
jaga	1	1	0	2	$\text{Log}(3/2) = 0,1760$
bbm	0	1	0	1	$\text{Log}(3/1) = 0,4771$
tidak	0	1	0	1	$\text{Log}(3/1) = 0,4771$
roket	0	1	0	1	$\text{Log}(3/1) = 0,4771$
ayo	0	0	1	1	$\text{Log}(3/1) = 0,4771$
investasi	0	0	1	1	$\text{Log}(3/1) = 0,4771$
kripto	0	0	1	1	$\text{Log}(3/1) = 0,4771$
peluang	0	0	1	1	$\text{Log}(3/1) = 0,4771$
Jumlah					5,123

Proses perhitungan TF dan IDF sudah didapat, kemudian mencari nilai bobot *tf-idf* sesuai dengan sentimen pada setiap dokumen, untuk penelitian ini untuk sentimen yang digunakan yaitu positif, negatif dan netral sesuai opini masyarakat. Perhitungan dapat dilihat pada Tabel 6, 7, 8.

Tabel 6 Hasil TF-IDF Label Positif

Term	TF	IDF	TF-IDF=TF*IDF
alhamdulillah	1	0,4771	0,4771
inflasi	1	0,1760	0,1760
indonesia	1	0,4771	0,4771
turun	1	0,4771	0,4771
jaga	1	0,1760	0,1760
Jumlah			1,7833

Tabel 7 Hasil TF-IDF Label Negatif

Term	TF	IDF	TF-IDF=TF*IDF
bbm	1	0,4771	0,4771
tidak	1	0,4771	0,4771
jaga	1	0,1760	0,1760

inflasi	1	0,4771	0,4771
roket	1	0,4771	0,4771
Jumlah			2,0844

Tabel 8 Hasil *TF-IDF* Label Netral

Term	TF	IDF	TF-IDF=TF*IDF
ayo	1	0,4771	0,4771
investasi	1	0,4771	0,4771
kripto	1	0,4771	0,4771
pluang	1	0,4771	0,4771
Jumlah			1,9084

3.5 Klasifikasi Naïve Bayes

Proses *tf-idf* telah didapat, selanjutnya perhitungan *naïve bayes* model *multinomialnb* dengan data uji, berikut tweet yang akan diuji dengan model klasifikasi *naïve bayes* dengan model yang digunakan *multinomialNB*, contoh data uji pada Tabel 9.

Tabel 9 Data *test*

Data uji	Hasil preprocessing	Sentimen
Harga kebutuhan naik, inflasi meroket	Harga butuh naik, inflasi roket	?

Proses pertama dilakukan menghitung *prior probability* pada setiap masing-masing kelas sentimen, perhitungan dan hasil pada Tabel 10 menggunakan persamaan (5).

Tabel 10 Hasil Probabilitas Setiap Sentimen

Rumus	Hasil
P(positif) = 1/3	0,3
P(netral) = 1/3	0,3
P(negatif) =1/3	0,3

Prior probability didapat, kemudian menghitung kondisi probabilitas untuk setiap masing-masing kata dan masing-masing sentimen pada contoh data latih. Perhitungan dapat dilihat pada Tabel 11,12,13 menggunakan persamaan (6).

Tabel 11 Perhitungan kondisi probabilitas kelas positif

Positif	Hasil
harga	$(0 + 1) / (1,7833 + 5,123) = 1 / 6,9063 = 0,144795332$
butuh	$(0 + 1) / (1,7833 + 5,123) = 1 / 6,9063 = 0,144795332$
naik	$(0 + 1) / (1,7833 + 5,123) = 1 / 6,9063 = 0,144795332$
inflasi	$(0,1760 + 1) / (1,7833 + 5,123) = 1,1760 / 6,9063 = 0,17027931$
roket	$(0 + 1) / (1,7833 + 5,123) = 1 / 6,9063 = 0,144795332$

Tabel 12 Perhitungan kondisi probabilitas kelas negatif

Negatif	Hasil
harga	$(0 + 1) / (2,0844 + 5,123) = 1 / 7,2074 = 0,138746289$
butuh	$(0 + 1) / (2,0844 + 5,123) = 1 / 7,2074 = 0,138746289$
naik	$(0 + 1) / (2,0844 + 5,123) = 1 / 7,2074 = 0,138746289$
inflasi	$(0,4771 + 1) / (2,0844 + 5,123) = 1,4771 / 7,2074 = 0,204942143$
roket	$(0,4771 + 1) / (2,0844 + 5,123) = 1,4771 / 7,2074 = 0,204942143$

Tabel 13 Perhitungan kondisi probabilitas kelas netral

Netral	Hasil
harga	$(0 + 1) / (1,9084 + 5,123) = 1 / 7,0314 = 0,142219188$
butuh	$(0 + 1) / (1,9084 + 5,123) = 1 / 7,0314 = 0,142219188$
naik	$(0 + 1) / (1,9084 + 5,123) = 1 / 7,0314 = 0,142219188$
inflasi	$(0 + 1) / (1,9084 + 5,123) = 1 / 7,0314 = 0,142219188$
roket	$(0 + 1) / (1,9084 + 5,123) = 1 / 7,0314 = 0,142219188$

Nilai *prior probability* dan kondisi probabilitas sudah ditemukan selanjutnya menghitung keseluruhan probabilitas atau *posterior probability* setiap sentiment dan kata. Perhitungan dapat dilihat pada Tabel 14 menggunakan persamaan (4).

Tabel 14 Hasil akhir *posterior probability*

Keterangan	Rumus	Hasil
positif	$0,3 \times 0,144795332 \times 0,144795332 \times 0,144795332 \times 0,17027931 \times 0,144795332$	2,24543
negatif	$0,3 \times 0,138746289 \times 0,138746289 \times 0,138746289 \times 0,204942143 \times 0,204942143$	3,36549
netral	$0,3 \times 0,142219188 \times 0,142219188 \times 0,142219188 \times 0,142219188 \times 0,142219188$	1,74547

Proses klasifikasi *naive bayes multinomialnb* bobot probabilitas data uji sudah didapat, kemudian menentukan nilai probabilitas terbaik dan didapatkan kelas negatif mendapatkan nilai probabilitas tertinggi yaitu 3,36549, maka data uji tersebut masuk kedalam kelas negatif.

3.6 Evaluasi

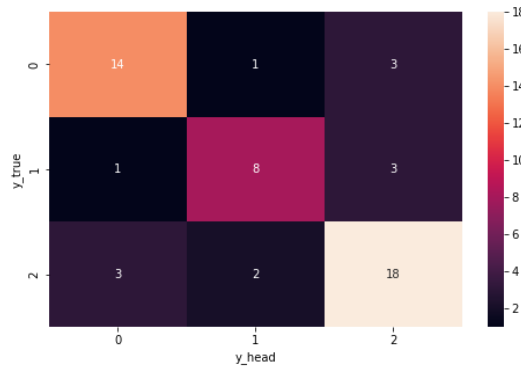
Evaluasi digunakan setelah proses pengujian model sudah dilakukan dan mendapatkan hasil berupa nilai *accuracy*, *precision*, *recall* dan *f1-score*. Proses klasifikasi penelitian ini menggunakan beberapa skenario pembagian dataset untuk menghasilkan model terbaik yaitu menggunakan 90:10, 80:20, 70:30 dalam pembagian data kemudian menggunakan *naive bayes* model *multinomialNB* sebagai proses pengklasifikasian analisis sentimen, skenario pembagian data tersebut kemudian diambil hasil evaluasi yang terbaik, berikut hasil evaluasi yang berdasarkan pembagian data pada Tabel 15.

Tabel 15 *Persentase Splitting Data dan Hasil*

Persentase Data		Jumlah Data		Hasil			
Data Train	Data Test	Data Train	Data Test	<i>Accuracy</i>	<i>precision</i>	<i>recall</i>	<i>F1-score</i>
90%	10%	473	53	75%	75%	74%	75%
80%	20%	420	106	63%	65%	64%	62%
70%	30%	158	368	61%	69%	64%	58%

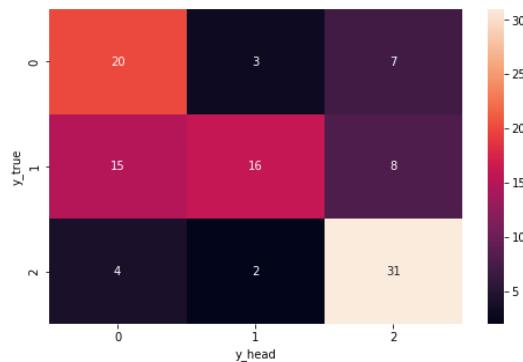
Dari hasil evaluasi tersebut hasil klasifikasi *naive bayes* dengan model *multinomialNB* untuk sekenario pembagian data 90% data train dan 10% data test menjadi model terbaik dengan

precision 75%, *recall* 74%, *f1-score* 75% dan *accuracy* 75%, kemudian dari hasil evaluasi tersebut menghasilkan *confusion matrix* sesuai dengan model yang terbaik, berikut gambar visualisasi pada Gambar 3.



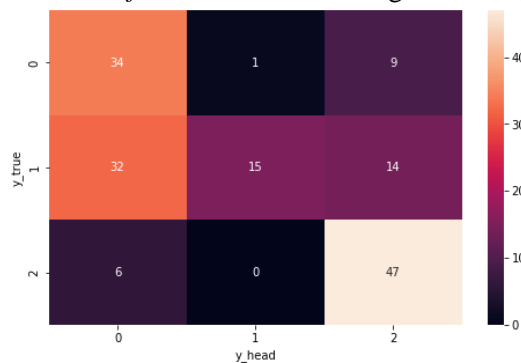
Gambar 3 *Confusion matrix* 90:10

Evaluasi kedua untuk skenario pembagian data 80% untuk data train dan 20% data test mendapatkan hasil *precision* 65%, *recall* 64%, *f1-score* 62% dan *accuracy* 63%, kemudian dari hasil evaluasi tersebut menghasilkan *confusion matrix*, berikut gambar *confusion matrix* pada Gambar 4.



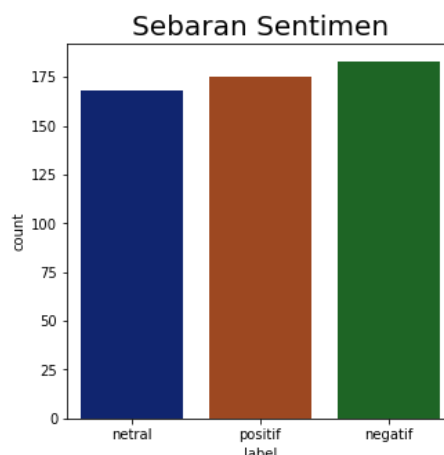
Gambar 4 *Confusion matrix* 80:20

Evaluasi ketiga untuk skenario pembagian data 70% data train 30% data test mendapatkan hasil *precision* 69%, *recall* 64%, *f1-score* 58% dan *accuracy* 61%, kemudian dari hasil evaluasi tersebut menghasilkan *confusion matrix*, berikut gambar *confusion* pada Gambar 5.



Gambar 5 *Confusion matrix* 70:30

Pembagian data tersebut berisi beberapa sentimen pada setiap tweet yaitu sentimen positif, negatif dan netral dengan jumlah masing-masing *tweet* ber sentimen positif 175, negatif 183 dan netral 168. Maka menghasilkan visualisasi seperti pada Gambar 6.



Gambar 6 Visualisasi Sebaran Sentimen

4. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian algoritma *naïve bayes classifier* dengan model *multinomialNB* untuk analisis sentimen terhadap inflasi di Indonesia.

1. Dataset yang digunakan sebanyak 1725 tweet yang kemudian dilakukan *preprocessing* menghasilkan data sebanyak 526 *tweet*.
2. Penelitian ini menghasilkan akurasi sebesar 75% dengan menggunakan pembagian data 90 % data train dan 10 % data test sehingga sudah cukup baik dalam mengklasifikasi sentimen.
3. Perlu dilakukan penelitian lebih lanjut yaitu dengan mencoba algoritma-algoritma klasifikasi lain agar menghasilkan sistem yang lebih baik
4. Perlu menambahkan varian jumlah data, karena semakin banyak dataset maka variasi kalimat juga semakin banyak, maka dalam mengklasifikasi lebih optimal.

DAFTAR PUSTAKA

- [1] D. Widiarsih and R. Romanda, "Analisis Faktor-Faktor yang Mempengaruhi Inflasi di Indonesia Tahun 2015-2019 dengan Pendekatan Error Corection Model (ECM)," *J. Akunt. dan Ekon.*, vol. 10, no. 1, pp. 119–128, 2020, doi: 10.37859/jae.v10i1.1917.
- [2] P. L. Harahap and Nurbaiti, "Jurnal Ilmu Kompuer, Ekonomi dan Manajemen," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 2, no. 2, pp. 3414–3420, 2022.
- [3] J. Matematika *et al.*, "Pengaruh Kenaikan BBM Terhadap Laju Inflasi di Indonesia," vol. 2, no. 1, pp. 25–32, 2023.
- [4] N. P. R. Yuliantini and K. D. Pramita, "Jurnal komunikasi hukum," *J. Komun. Hukum, Volume 7 Nomor 1 Februari 2021*, vol. 8, no. 1, pp. 469–480, 2022, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/jkh/issue/view/863>
- [5] lyla rahma Adyani and R. djoko Sampurno, "Analisis faktor-faktor yang mempengaruhi yang mempengaruhi Profitabilitas," *J. Akunt. dan Keuang.*, vol. 2, no. 2, pp. 40–57, 2019.
- [6] A. Setiawan and I. Mbae, "Analisis Perbandingan Harga Minyak Mentah Dunia Sebelum dan Semasa Perang Rusia-Ukraina : Event Study," *Ekomen J. Ilm. Ekomen*, vol. 22, no. 2, pp. 20–28, 2022.
- [7] R. Hastuti, I. Irawan, and A. Hukum, "Pengaruh Inflasi , Nilai Tukar , Suku Bunga dan Produk Domestik Bruto terhadap Return Saham pada Perusahaan Manufaktur (The Effect of Inflation , Exchange Rate , Interest Rate and Gross Domestic Products on Stock Returns in Manufacturing Companies)," vol. 2, no. 1, pp. 21–36, 2023.
- [8] Y. Sari, E. Winarni, and M. Amali, "Analisis Hubungan Pertumbuhan Ekonomi, Angkatan Kerja dan Inflasi di Indonesia: Pendekatan Vector Error Correction Model (VECM),"

- Ekon. J. Econ. Bus.*, vol. 7, no. 1, p. 627, 2023, doi: 10.33087/ekonomis.v7i1.821.
- [9] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- [10] A. L. D. tavares Duarte and Eddy Nurraharjo, "Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 1–8, 2023, doi: 10.36595/jire.v6i1.779.
- [11] Z. N. Aulia, G. K. Jati, and I. Santoso, "ANALISIS SENTIMEN TANGGAPAN PUBLIC MENGENAI E-TILANG MELALUI MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA NAIVE BAYES," vol. 7, no. 2.
- [12] R. Puspitasari, Y. Findawati, M. A. Rosid, P. S. Informatika, and U. M. Sidoarjo, "Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Analisis Sentimen Terhadap Inflasi Pasca Covid-19 Berdasarkan Twitter Dengan Metode Klasifikasi K-Nearest Neighbor Dan," vol. 4, no. 4, pp. 1–11, 2023, doi: 10.20884/jutif.
- [13] P. D. Wulandari, U. Enri, and A. Primajaya, "Analisis Sentimen Terhadap Kenaikan Iuran Bpjs Kesehatan Pada Twitter Menggunakan Naive Bayes Classifier," *J. Ilmu Komput. dan Teknol.*, vol. 1, no. 2, pp. 18–22, 2020, doi: 10.35960/ikomti.v1i2.575.
- [14] S. Mujahidin, B. Prasetio, and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naive bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.)*, vol. 10, no. 3, p. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [15] S. Mandasari, B. H. Hayadi, and R. Gunawan, "Sentiment Analysis of Online Transportation Users Towards Grab Indonesia Services Using Multinomial Naive Bayes Classifier," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, p. 118, 2022, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/article/view/5635>
- [16] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.
- [17] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [18] P. G. Aryanti and I. Santoso, "ANALISIS SENTIMEN PADA TWITTER TERHADAP MOBIL LISTRIK MENGGUNAKAN ALGORITMA NAIVE BAYES," vol. 7, no. 2, pp. 133–137, 2023.
- [19] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 2, pp. 195–200, 2022, doi: 10.32736/sisfokom.v11i2.1412.
- [20] A. M. Siregar *et al.*, "ANALISIS SENTIMEN MASYARAKAT TERHADAP UNIVERSITAS BUANA PERJUANGAN KARAWANG DENGAN AIGORITME SVM dan NAIVE BAYES," pp. 25–36, 2023.
- [21] F. Septianingrum, J. H. Jaman, and U. Enri, "Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1431, 2021, doi: 10.30865/mib.v5i4.3260.