

# ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME NAÏVE BAYES

Anggi Ayu Ningtyas<sup>1\*</sup>, Achmad Solichin<sup>2</sup>, Rizky Pradana<sup>3</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

<sup>3</sup>Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur

Email: <sup>1\*</sup>1911501193@student.budiluhur.ac.id, <sup>2</sup>achmad.solichin@budiluhur.ac.id,

<sup>3</sup>rizky.pradana@budiluhur.ac.id

(\* : corresponding author)

(Naskah masuk: 6 Maret 2023, diterima untuk diterbitkan: 12 Maret 2023)

## Abstrak

YouTube adalah situs berbagi media, yaitu semacam hiburan virtual untuk berbagi media video dan suara. YouTube telah menjadi salah satu platform tontonan video paling populer saat ini. Beberapa waktu lalu, Youtube Indonesia disibukkan dengan unggahan ke channel Raymond Chin pada 3 Oktober 2022 tentang prediksi ekonomi tahun 2023. Oleh karena itu, akan sangat sulit untuk memahami apakah netizen yang menggunakan situs youTube umumnya positif negatif. atau netral. Dengan video yang ditransfer ini, diharapkan sebuah metode untuk menginvestigasi komentar-komentar yang begitu banyak sehingga komentar-komentar tersebut dapat bernilai dan dapat ditarik tujuan yang signifikan. Pelonggaran perputaran keuangan ini dapat berlangsung dalam jangka waktu yang sangat lama, bahkan bertahun-tahun. Sementara itu, penghentian moneter global, pembenahan strategi moneter, tekanan internasional, perselisihan pasar keuangan hingga perubahan lingkungan penyebab menjadi faktor yang mempengaruhi elemen ekonomi pada tahun 2023. Oleh karena itu, untuk mengetahui sentimen tentang perkiraan dan penurunan ekonomi pada tahun 2023 membutuhkan kerangka kerja untuk memutuskan apakah sentimen publik "positif", "negatif", atau "netral". Naïve Bayes adalah strategi urutan yang ditetapkan dalam hipotesis Bayes. khususnya mengantisipasi masa depan mengingat pengalaman masa lalu. pengklasifikasian Naive Bayes ini menjalankan metode karakterisasi item terarah di masa mendatang dengan membagikan tanda kelas ke kejadian yang memanfaatkan probabilitas. Dalam ulasan ini, pengaturan akan diselesaikan menggunakan strategi Naïve Bayes dengan 500 data Untuk pengujian gunakan 10 Folds Cross Validation dengan mempartisi informasi persiapan menjadi beberapa bagian untuk pengujian mendapatkan keakuratan yang berbeda karena menggunakan informasi persiapan yang tidak teratur. Setelah pengujian menghasilkan "positif" bernilai 40%, "negatif" 26% dan "netral" 34%. Konsekuensi dari pengujian ini menunjukkan bahwa dapat membedakan realitas data komentar youtube dengan menggunakan strategi Naïve Bayes.

**Kata kunci:** *resesi, youtube, sentimen masyarakat, naïve bayes, text mining*

## ***SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON PREDICTION OF ECONOMIC RECESSION IN 2023 USING THE NAÏVE BAYES ALGORITHM***

### ***Abstract***

*YouTube is a media sharing site, a kind of virtual entertainment for sharing video and sound media. YouTube has become one of the most popular video viewing platforms today. Some time ago, Youtube Indonesia was busy with uploads to Raymond Chin's channel on October 3, 2022 regarding economic predictions for 2023. Therefore, it will be very difficult to understand whether netizens who use the YouTube site are generally positive or negative. or neutral. With this transferred video, it is hoped that a method for investigating the large number of comments so that these comments can be of value and can be drawn a significant purpose. This easing of financial cycles can last for a very long period of time, even years. Meanwhile, global monetary shutdowns, monetary strategy reforms, international pressures, financial market disputes to environmental changes are the factors that affect the economic elements in 2023. Therefore, to know sentiments about forecasts and economic downturn in 2023 requires a framework for decide whether public sentiment is "positive", "negative" or "neutral". Naïve Bayes is a sequence strategy defined in the Bayesian hypothesis. especially anticipating the future given past experiences.*

*This Naive Bayes classifier implements a future directed item characterization method by assigning class marks to events that take advantage of probabilities. In this review, the arrangement will be completed using the Naive Bayes strategy with 500 data. For testing, use 10 Folds Cross Validation by partitioning the preparation information into several parts for testing to obtain different accuracy because it uses irregular preparation information. After testing it produces a "positive" value of 40%, "negative" 26% and "netral" 34%. The consequences of this test show that it can distinguish the reality of YouTube comment data using the Naive Bayes strategy.*

**Keywords:** recession, youtube, community sentiment, naive bayes, text mining

## 1. PENDAHULUAN

Sebuah studi dari *We are Social* menunjukkan bahwa ada 150 juta orang Indonesia yang secara efektif memanfaatkan hiburan *online*. Dari sekian banyak hiburan virtual yang dimanfaatkan, yang paling dinamis dan dominan adalah YouTube dengan 88%. Banyaknya pelanggan terjadi karena substansinya semakin berbeda dan berkembang. Diperkirakan lebih dari 600 pengguna di Indonesia memiliki lebih dari 1 juta endorser [1].

Belum lama ini, YouTube Indonesia sempat disibukkan oleh unggahan *channel* Raymond Chin [2] pada 3 Oktober 2022 tentang prediksi ekonomi tahun 2023. Penurunan ekonomi atau resesi adalah istilah yang digunakan untuk menggambarkan apa yang terjadi di mana siklus moneter suatu negara menjadi lambat atau memburuk. Pelonggaran perputaran keuangan ini dapat berlangsung dalam rentang waktu yang sangat lama, bahkan bertahun-tahun, karena pertumbuhan PDB (Produk Domestik Bruto) suatu negara menurun selama dua kuartal dan berlangsung terus menerus.

Oleh karena itu, akan sangat sulit untuk memahami apakah *netizen* yang menggunakan situs YouTube biasanya bersikap positif, negatif, atau netral. Dengan video yang ditransfer ini, diharapkan ada prosedur untuk menginvestigasi komentar-komentar yang sangat banyak sehingga komentar-komentar tersebut dapat bernilai dan tujuan yang signifikan dapat ditarik.

Penelitian terkait analisis media sosial, khususnya yang terkait data komentar Youtube sudah banyak dilakukan. Beberapa penelitian terkini yang cukup menarik salah satunya oleh Sanjaya dan Lakhsamana yang menganalisis data komentar Youtube terkait susunan Kabinet Indonesia Maju [3]. Data dikumpulkan dari youtube *comment* di 3 *channel* youtube yaitu SCTV, CNN, dan Kompas TV dengan total data komentar yang didapat sebanyak 7.366 data komentar. Pengujian dengan menggunakan *setiwordnet* dan sentimen *lexicon* indonesia pada tahap fitur ekstarki ditemukan bahwa percobaan dengan menggunakan sentiment *lexicon* Indonesia lebih baik dari SentiWordNet.

Selanjutnya, pada penelitian lainnya oleh Dona dkk. [4] dilakukan analisis komentar Youtube menggunakan 2 (dua) algoritma, yaitu KNN dan Naive Bayes. Hasilnya algoritma KNN unggul dibandingkan Naive Bayes. Namun pada penelitian lainnya yang lebih terkini, algoritma Naive Bayes

memiliki performa yang baik dengan akurasi hingga 90% [5]. Data tersebut diambil sebanyak 1400 komentar dalam waktu satu menit. Komentar diambil pada tanggal 8 juli 2021. Jumlah data komentar sebanyak 4.229 terdiri dari 1023 data positif, 1030 data negatif dan 2.176 data bukan bahasa Indonesia. Penelitian yang berbeda menggunakan algoritma Naive Bayes untuk menganalisis sentimen komentar Youtube pada *channel* *Food Vlogger* Berbahasa Indonesia [6]. Analisis terhadap 1702 komentar menghasilkan komentar bersentimen positif sebanyak 848 data dan negatif sebanyak 854 data. Pengujian dengan komposisi data training dan testing 70:30 menghasilkan nilai akurasi sebesar 90,60%. Metode Naive Bayes memang banyak digunakan dalam berbagai penelitian analisis sentimen pada data media sosial, seperti analisis sentimen terhadap manajemen bencana [7], penyelenggaraan MotoGP [8], komentar pada grup Facebook [9], dan penyelenggaraan Pilkada [10]. Selain itu, metode SVM juga cukup populer digunakan seperti untuk analisis sentimen terkait transportasi umum [11] dan kebijakan PSBB [12].

Informasi diambil dari komentar YouTube yang dibuat pada 11 November 2022, Informasi yang digunakan hanya 500 data dan Sistem hanya mengetahui opini tentang unggahan saluran di Youtube bernilai positif, netral atau negatif. Tujuan yang ingin dicapai dalam eksplorasi ini adalah membuat kerangka kerja untuk melaksanakan strategi Naive Bayes dalam kerangka penemuan opini sentimen masyarakat dalam unggahan video youtube di saluran Raymond Chin tentang prediksi ekonomi berjudul 2023 : Menuju Kehancuran Dunia. dengan hasil apakah unggahan mendapat pendapat yang "positif", "negatif" atau "netral".

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Penelitian ini menggunakan API Library yaitu Youtube Data API v3 dan APIs & Services untuk meminta API Key dari akun Google Developers untuk mengakses Google *Developer Console*. Google *Developers Console* sendiri mempunyai tools APIs & Services dimana kita diharuskan *Enabled APIs & Services* yang akan *Redirect* kedalam API Library dengan *Search Youtube Data API v3* dan *Create Credentials* untuk mendapatkan API Key yang akan dipakai untuk *crawling* data pengumpulan dataset ini

menggunakan sentimen komentar video youtube pada tanggal 14 November 2022 sebanyak 500 data.

### 2.2 Labelling Data

Pada tahapan manual labeling peneliti menggunakan 4 pakar untuk melabelkan data. Penulis menggunakan pakar yang berasal dari jenjang pendidikan S1 (Sarjana). sebanyak 2 pakar berasal dari bidang ilmu sastra Indonesia dan dua pakar berasal dari bidang ilmu ekonomi. alasan penulis memilih pakar jurusan bidang ilmu sastra Indonesia dikarenakan dapat menganalisa bahasa Indonesia dengan baik dan mahir dalam kemampuan tata bahasa yang relevan dengan penelitian ini selanjutnya bidang ilmu ekonomi dikarenakan penelitian ini tentang analisa prediksi resesi ekonomi dibutuhkan sudut pandang dari bidang ilmu ekonomi. Pada proses ini responden diminta untuk menentukan apakah data komentar youtube tersebut bernilai positif, negatif atau netral Setelah keempat responden diminta melabelkan, penulis mengambil sentimen terbanyak untuk menentukan hasil dari pelabelan. Jika hasil sentimen seimbang maka penulis menjadi *decision maker* (penentu) untuk menentukan teks komentar tersebut cenderung bernilai positif, negatif atau netral. contohnya seperti tabel berikut:

Tabel 1. Pakar

Pakar 1	Pakar 2	Pakar 3	Pakar 4	Hasil
Negatif	Negatif	Negatif	Negatif	Negatif
Netral	Positif	Positif	Negatif	Positif
Negatif	Negatif	Netral	Netral	Netral

### 2.3 Preprocessing

*Preprocessing* merupakan tahapan proses untuk membersihkan data dari kata-kata atau komentar youtube yang tidak di perlukan serta kata-kata yang tidak memiliki makna. Proses ini dilakukan sesuai dengan isi data dari proses pengambilan data atau *crawling* data youtube. Adapun tahapannya *tokenizing*, *case folding*, *stopword removal* dan *stemming*.

#### 2.3.1 Tokennizing

Tahapan pertama yang dilakukan yaitu *Tokenizing* metode untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan untuk proses analisis teks lebih lanjut. mencakup proses menghapus angka dan tanda baca yang tidak perlu, dan *whitespace*.

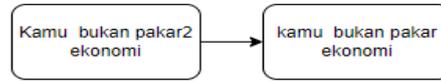


Gambar 1. Tokennizing

#### 2.3.2 Case Folding

Langkah selanjutnya yang umumnya dilakukan adalah tahap *case folding*. Data yang diberikan tidak selalu konsisten dalam penggunaan huruf kapital dan tidak terstruktur. *Case folding* berperan dalam

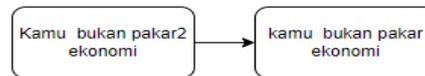
menyamakan penggunaan huruf kapital. Sebagai contoh, jika kita memiliki data teks "Ayu NinG", maka dengan melakukan *case folding* kita akan mengubah semua huruf menjadi huruf kecil (*lowercase*). Karakter-karakter lain yang tidak termasuk huruf atau angka, seperti tanda baca dan spasi, dianggap sebagai delimiter.



Gambar 2. Case Folding

#### 2.3.3 Stopword Removal

Tahap penyaringan menghilangkan kata-kata kurang penting dari hasil token menggunakan stoplist (yang menghilangkan kata-kata yang kurang penting) atau daftar kata (yang menyimpan kata-kata penting). *Stopwords* adalah kata-kata yang sering digunakan tetapi dianggap tidak berarti. *Stopwords* dalam bahasa Indonesia menggabungkan "yang", "dan", "di, dll. *Stopwords* digunakan untuk menghilangkan kata-kata dalam teks yang tidak banyak bicara sehingga kita hanya bisa fokus pada yang penting. *Stopwords* dapat digunakan dalam sebuah berbagai konteks, termasuk mesin telusur. Jika "kamu bukan ekonomi" adalah yang kami cari, sistem penelusuran seharusnya memberi bobot lebih pada dokumen tentang "memahami seorang ekonomi" daripada dokumen tentang "kamu bukan ekonomi". Ini dapat dicapai dengan menghilangkan analisis kata dari daftar *stopwords*.



Gambar 3. Stopword

#### 2.3.4 Stemming

*Stemming* adalah proses mengubah infleksi suatu kata menjadi bentuk dasarnya, tetapi makna bentuk dasarnya berbeda dengan kata dasarnya. Biasanya, saat kita menelusuri dokumen "kamu bukan pakar ekonomi" kita ingin mencocokkan semua kemungkinan ejaan untuk menemukan ejaan yang relevan. Untuk bekerja pada sifat data yang diperoleh, pemanfaatan aturan imbuhan atau kamus kata dasar untuk meningkatkan kualitas informasi yang diperoleh merupakan salah satu cara untuk melakukan *stemming*. Menggunakan aturan infleksi atau kamus kata dasar dapat digunakan untuk melakukan *stemming*.



Gambar 4. Stemming

### 2.4 Pembobotan Term Frequency

Tahapan pembobotan (*term frequency*) dilakukan setelah data sudah bersih (*preprocessing*) dihitung jumlah kemunculan kata yang terdapat dalam komentar youtube pada data latih (training) Cara yang

paling tidak rumit untuk menangani ide ini adalah mengomunikasikan beratnya kata t saat kata itu muncul dalam laporan d. Misalnya, jika dalam catatan kata pemalsuan muncul berkali-kali, nilai TF adalah 10. Konsep term *frequency* melihat catatan sebagai sekarang kata di mana permintaan peristiwa kata diabaikan dan hanya jumlah peristiwa kata yang penting [13]. Nilai TF ditentukan menggunakan Persamaan (1).

$$TF = \frac{\text{Jumlah Frekuensi Kata Terpilih}}{\text{Jumlah Kata}} \dots\dots\dots (1)$$

Contoh kalimat = “kamu bukan pakar ekonomi”

berikut perhitungan kemunculan kata dapat dilihat pada tabel berikut :

Term	Count
kamu	1
bukan	1
pakar	1
ekonomi	1

**2.5 Naïve Bayes**

Pada tahap akhir setelah melakukan pembobotan, penulis melakukan pengujian klasifikasi dengan metode naïve bayes, setelah setiap kata dihitung dan diketahui nilai pembobotannya yang dilakukan pada proses pembobotan term *frequency* (tf) setelah itu data tersimpan pada variabel dan didapatkan frekuensi kata pada keseluruhan dokumen latih (*training*), selanjutnya data tersebut dimasukkan ke dalam pengujian model atau metode kalsifikasi naïve bayes untuk menghitung nilai probabilitasnya. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter. Naïve Bayes dapat dirumuskan sebagai berikut:

$$P(A|B) = P(B|A)P(A)P(B) \dots\dots\dots (2)$$

Keterangan :

- P(A | B) : *Probabilitas* A terjadi dengan bukti bahwa B telah terjadi (*probabilitas superior*)
- P(B | A) : *Probabilitas* B terjadi dengan bukti bahwa A telah terjadi
- P(A) : Peluang terjadinya A
- P(B) : Peluang terjadinya B

**2.6 10 Fold Cross Validation**

Pada tahapan ini dilakukan pengujian 10 *Fold-Cross Validation*. Untuk mencari mendapatkan akurasi pada klasifikasi naïve bayes dengan mengambil data latih bersih (*preprocessing*). Selanjutnya data dipisahkan secara acak dan dikelompokkan sebanyak 10 kali kemudian data tersebut dibagi menjadi dua sub set yaitu data latih dan data uji, *Cross validation* akan menggunakan 9 *fold* untuk pelatihan dan 1 *fold* [13]

**Tabel 3. Confusion Matrix**

	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

- True Positive* (TP) :Jumlah data yang bernilai positif dan diprediksi benar sebagai positif.
- False Positive* (FP) :Jumlah data yang bernilai negatif tetapi diprediksi sebagai positif.
- False Negative* (FN) :Jumlah data yang bernilai positif tetapi diprediksi sebagai negatif.
- True Negative* (TN) : Jumlah data yang bernilai negatif dan diprediksi benar sebagai negatif.

**3. HASIL DAN PEMBAHASAN**

**3.1 Requirement Analysis**

Tahapan selanjutnya yaitu *requirement analysis* yaitu tahap interaksi intensif antara analis sistem dengan pemakai sistem (*end user*). Terdapat dua pemakai user (*end user*) dalam sistem ini yaitu pemilik *channel* atau *content creator* yang dimana sistem ini dapat melihat reaksi atau sentimen masyarakat terhadap video yang diunggah untuk dijadikan perbaikan dan kedua yaitu masyarakat yang berkomentar di unggahan video tersebut untuk menjadikan pembelajaran, dapat mengetahui sentimen dan ditarik kesimpulannya. Dalam tahapan ini peneliti menggunakan metode analisis kebutuhan yaitu *Document Survey* yang dimana peneliti melakukan pengamatan terhadap dokumen untuk menentukan kebutuhan analisa pengguna berikut penjelasannya:

- a) Bagaimana metode itu digunakan.  
Sistem ini menggunakan metode Naïve Bayes yang dimana klasifikasi tersebut mengklasifikasi berdasarkan probabilitas sederhana yang dirancang agar dapat digunakan dengan asumsi antar variabel independen. Cara kerja yaitu menghitung peluang dari satu kelas dari masing-masing kelompok atribut dan menentukan kelas mana yang paling optimal atau dapat dilakukan berdasarkan kategori yang pengguna masukkan pada perangkat lunak. Untuk lebih jelasnya dapat dilihat pada tahapan naïve bayes.
- b) Target dari metode.  
Sistem ini mempunyai tujuan atau target dalam mengimplementasikan metode yaitu dapat klasifikasikan kategorial sentimen yaitu positif, negatif, dan netral
- c) Keuntungan metode.

Pengimplementasian metode naïve bayes dalam sistem ini mempunyai keuntungan atau manfaat yaitu membutuhkan lebih sedikit data pelatihan dalam sistem ini hanya menggunakan 300 data latih untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

- d) Kerugian metode.  
Pengimplementasian metode naïve bayes dalam sistem ini mempunyai kelemahan yaitu keakuratan tidak bisa dihitung apabila menggunakan satu probabilitas. Untuk mengklasifikasikan metode ini

perlu ketiga kategorial agar dapat berkorelasi antar probabilitas positif, probabilitas negatif dan probabilitas netral untuk membuktikan akurasi.

e) Kapan metode tersebut baik digunakan.

Sistem ini baik digunakan apabila data latih yang digunakan terstruktur dan beraturan untuk meminimalisir kegagalan sistem dalam klasifikasi naïve bayes ini menggunakan kalimat yang tidak terlalu panjang dan tahapan pada *preprocessing* dapat bekerja secara optimal untuk menghasilkan kalimat yang baik sehingga melakukan klasifikasi bisa mendapatkan akurasi yang tinggi.

### 3.2 Pengumpulan Data

Tahap pengumpulan data merupakan tahapan pertama kali dilakukan. Data yang dikumpulkan pada penelitian ini bersumber dari sosial media youtube. Pada penelitian ini, data yang digunakan merupakan kata kunci *Resesi*. Proses pengambilan komentar youtube menggunakan *Google Collab* dengan memasukan *Api Key* dari *Google Developer*. Total komentar yang diperoleh sebanyak 500 data. Selanjutnya data yang berhasil disortir untuk menghasilkan data yang relevan selanjutnya data yang sesuai akan dikumpulkan dijadikan dataset.

### 3.3 Labelling

Tahapan melabelkan data ini dilakukan secara manual yaitu menggunakan responden dengan kriteria yang telah disepakati adapun kriteria nya yaitu :

1. Minimal mahasiswa/i dengan jurusan sastra indonesia & fakultas ekonomi
2. Mempunyai akun youtube
3. Mencari 4 responden

dengan kriteria tersebut peneliti mendapatkan empat responden, dua responden berasal dari jurusan sastra indonesia, satu responden berasal dari jurusan manajemen dan satu responden berasal dari jurusan akuntansi. Empat responden tersebut melabel komentar youtube tersebut bernilai positif, negatif atau netral. Selanjutnya dari keempat responden tersebut dihitung sentimen terbanyak, jika keempat responden tersebut memiliki pendapat berbeda atau ketika sentimen dihitung bernilai sama maka peneliti menjadi (*decision maker*)

### 3.4 Term Frequency

Tahapan pembobotan ini akan dihitung nilai kemunculan kata pada setiap dokumen

**Tabel 4.** *Term Frequency*

Komentar	Label
kamu bukan pakar ekonomi	negatif
kamu turun ekonomi zaman	positif
tutup akses wilayah	netral
dampak terap sistem pasar bebas ajar terap sistem islam	positif
tidak ambil pusing orang desa mau gelap mau tidak aman	netral
tenang kita yakin atur dunia	netral
allah kasih sayang tidak khawatir	netral

pembobotan berdasarkan klasifikasi atau kategorial positif, negatif atau netral

**Tabel 5.** Pembobotan *Term Frequency*

Term	Positif	Negatif	Netral
kamu	1	1	0
bukan	1	0	0
pakar	1	0	0
ekonomi	1	0	0
zaman	1	0	0
tutup	1	0	0
akses	1	0	0
wilayah	1	0	0
tidak	2	0	0
ambil	1	0	0
pusing	1	0	0
orang	1	0	0
desa	1	0	0
mau	2	0	0
gelap	1	0	0
aman	1	0	0
tenang	0	0	1
kita	0	0	1
yakin	0	0	1
atur	0	0	1
dunia	0	0	1
allah	0	0	1
kasih	0	0	1
sayang	0	0	1
khawatir	0	0	1
dampak	0	0	1
terap	0	0	1
sistem	0	0	2
pasar	0	0	1
bebas	0	0	1
ajar	0	0	1
bukan	0	1	1
pakar	0	1	1
ekonomi	0	1	1

### 3.5 Naïve Bayes

Berikut tahapan *naïve bayes*

#### 3.5.1 Probabilitas Perkelas

Pada tahapan ini dilakukan dengan 5 data yang mempunyai kelas masing- masing yaitu 2 komentar bernilai positif, 2 komentar bernilai netral dan satu komentar bernilai negatif. Dengan jumlah kelas sebanyak 3 kategori yaitu “positif”, “negatif”, “netral” Berikut tabel nya:

**Tabel 6.** Label

Komentar	Label
kamu bukan pakar ekonomi	negatif
kamu turun ekonomi zaman tutup akses wilayah	positif
dampak terap sistem pasar bebas ajar terap sistem islam	netral

Komentar	Label
tidak ambil pusing dalam mau gelap mau tidak aman	positif
tenang kita yakin atur dunia allah kasih sayang tidak khawatir	netral

Selanjutnya, perhitungan *probabilitas* untuk kelas “Positif” *Probabilitas* disimbolkan sebagai p.

$$P(\text{Positif}) = \frac{\text{Jumlah Kelas Positif}}{\text{Jumlah Data Latih}} \dots\dots\dots (3)$$

$$P(\text{Positif}) = \frac{2}{5} = 0,4$$

*Probabilitas* untuk kelas “positif” dihitung dan memperoleh nilai yaitu 0,4.

Selanjutnya, perhitungan *probabilitas* untuk kelas “Negatif”

$$P(\text{Negatif}) = \frac{\text{Jumlah Kelas Negatif}}{\text{Jumlah Data Latih}} \dots\dots\dots (4)$$

$$P(\text{Negatif}) = \frac{1}{5} = 0,2$$

*Probabilitas* untuk kelas “negatif” dihitung dan memperoleh nilai yaitu 0,2

Selanjutnya, perhitungan *probabilitas* untuk kelas “Netral”

*Probabilitas* disimbolkan sebagai p.

$$P(\text{Netral}) = \frac{\text{Jumlah Kelas Negatif}}{\text{Jumlah Data Latih}} \dots\dots\dots (5)$$

$$P(\text{Netral}) = \frac{2}{5} = 0,4$$

*Probabilitas* untuk kelas “netral” dihitung dan memperoleh nilai yaitu 0,4.

Maka, disimpulkan untuk 5 data tersebut dengan 3 kategori ”positif”, ”negatif” dan ”netral”. Dari perhitungan tersebut “Positif” dan “Netral” menghasilkan *probabilitas* 0,4 untuk ”negatif” menghasilkan *probabilitas* 0,2

**3.5.2 Probabilitas Perkata**

Nilai *Frekuensi* Kata pada Tabel 7.

Proses perhitungan *probabilitas* kata label “positif”

- Menghitung *probabilitas* kata “kamu”
 
$$\frac{(100 + 1)}{(1700 + 500)} + \frac{101}{2200} = 0,04590909$$
- Menghitung *probabilitas* kata “bukan”
 
$$\frac{(30 + 1)}{(1700 + 500)} + \frac{31}{2200} = 0,01409090$$
- Menghitung *probabilitas* kata “pakar”
 
$$\frac{(20 + 1)}{(1700 + 500)} + \frac{21}{2200} = 0,00954545$$
- Menghitung *probabilitas* kata “ekonomi”

$$\frac{(15 + 1)}{(1700 + 500)} + \frac{16}{2200} = 0,00727272$$

**Tabel 7.** Frekuensi Kata

Kata	kamu	bukan	pakar	ekonomi
<i>Frekuensi</i> Positif	100	30	20	15
<i>Frekuensi</i> Negatif	120	40	7	5
<i>Frekuensi</i> Netral	80	70	30	10

Proses perhitungan *probabilitas* kata label “negatif”

- Menghitung *probabilitas* kata “kamu”
 
$$\frac{(120 + 1)}{(1400 + 500)} + \frac{121}{1900} = 0,06368421$$
- Menghitung *probabilitas* kata “bukan”
 
$$\frac{(40 + 1)}{(1700 + 500)} + \frac{41}{1900} = 0,02157894$$
- Menghitung *probabilitas* kata “pakar”
 
$$\frac{(7 + 1)}{(1700 + 500)} + \frac{8}{1900} = 0,00421052$$
- Menghitung *probabilitas* kata “ekonomi”
 
$$\frac{(5 + 1)}{(1700 + 500)} + \frac{6}{1900} = 0,00315789$$

Proses perhitungan *probabilitas* kata label “netral”

- Menghitung *probabilitas* kata “kamu”
 
$$\frac{(80 + 1)}{(2000 + 500)} + \frac{81}{2500} = 0,0324$$
- Menghitung *probabilitas* kata “bukan”
 
$$\frac{(70 + 1)}{(2000 + 500)} + \frac{71}{2500} = 0,0284$$
- Menghitung *probabilitas* kata “pakar”
 
$$\frac{(30 + 1)}{(2000 + 500)} + \frac{31}{2500} = 0,0124$$
- Menghitung *probabilitas* kata “ekonomi”
 
$$\frac{(10 + 1)}{(2000 + 500)} + \frac{11}{2500} = 0,0044$$

**3.6 Implementasi Aplikasi**

Pada tahapan ini menjelaskan tentang *aplikasi* yang telah dibuat

**3.6.1 Menu Data Training dan Testing**

Pada menu data *training* berfungsi untuk menampung data latih atau *training* dan menu data *testing* berfungsi untuk menampung data uji atau *testing* yang telah dilakukan *labelling* terdapat icon penunjuk lainnya.



Gambar 5. Training

### 3.6.2 Menu Preprocessing Training dan Testing

Pada menu ini berfungsi untuk membersihkan data *training* yang telah di *labelling* dengan tahapan *preprocessing* yaitu *tokenisasi*, *case folding*, *stopword* dan *stemming*. terdapat *icon* penunjang lainnya



Gambar 6. Preprocessing Training

### 3.6.3 Menu Pengujian Akurasi

Pada menu pengujian *confusion matrix* digunakan untuk menghitung nilai *akurasi*, *presicion*, dan *recall*. Keempat metode *evaluasi* tersebut sangat bermanfaat untuk mengukur *performa* dari *classifier*. Berikut

	TP	FP	FN	F1	ACCURACY	PRECISION	RECALL
Positif	120	10	20	0.857	0.857	0.857	0.857
Negatif	10	15	5	0.667	0.667	0.667	0.667
Akurasi					0.762		
Presision of Positif					0.857		
Presision of Negatif					0.667		
Recall of Positif					0.857		
Recall of Negatif					0.667		
Akurasi (F1)					0.762		

Gambar 7. Pengujian Akurasi Confusion Matrix

### 3.6.4 Menu Pengujian Naïve Bayes

Pada menu pengujian ini setelah melakukan pengujian *confusion matrix* maka tahap berikutnya yaitu klasifikasi *naïve bayes* terdapat hasil nilai *probabilitas* pada setiap kategori yaitu "positif", "negatif" dan "netral"

Kategori	Probabilitas	Jumlah
Positif	0.857	120
Negatif	0.667	25
Netral	0.762	100

Gambar 8. Pengujian Naive Bayes

### 3.6.5 Menu Visualisasi Hasil

Pada menu ini berfungsi menu untuk menampilkan grafik hasil dari pengujian dan pemodelan



Gambar 9. Visualisasi Hasil

## 4. KESIMPULAN

Berdasarkan hasil evaluasi dari aplikasi sentimen komentar *youtube*. Maka dapat disimpulkan bahwa:

Dari hasil klasifikasi maka dapat disimpulkan bahwa sentimen masyarakat terhadap prediksi *resesi 2023* bernilai positif, Sistem dapat menganalisa sentimen masyarakat bernilai positif, negatif, dan netral, dalam ulasan ini menggunakan metode *naïve bayes* dengan menggunakan 500 data yang telah diujikan menggunakan *fold cross validation* dengan 10 putaran iterasi kemudian data tersebut dibagi atau *split* menjadi dua *subset* yaitu data latih dan data uji yang dimana 9 *fold* untuk pelatihan dan 1 *fold* untuk pengujian maka menghasilkan nilai 40 % untuk kelas "positif", 26% untuk kelas "negatif" dan 34% untuk kelas "netral".

## DAFTAR PUSTAKA

- [1] F. F. Abdulloh And I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19," *Csrid (Computer Science Research And Its Development Journal)*, Vol. 13, No. 3, P. 141, Nov. 2021, Doi: 10.22303/Csrid.13.3.2021.141-148.
- [2] Raymond Chin, "Raymond Chin Youtube Channel," *Youtube*. <https://www.youtube.com/@Raymondchins> (Accessed Feb. 18, 2023).
- [3] G. Sanjaya And K. Muslim Lhaksana, "Analisis Sentimen Komentar Youtube Tentang Terpilihnya Menteri Kabinet Indonesia Maju Menggunakan Lexicon Based," *Eproceedings Of Engineering*, Vol. 7, No. 3, 2020.
- [4] F. Dona, I. Maulida, And B. Nugraha, "Klasifikasi Dan Analisis Dataset Komentar Video Youtube Menggunakan Supervised Learning," In *Seminastika*, Nov. 2021, Pp. 86-90. Doi: 10.47002/Seminastika.V3i1.232.
- [5] M. Hudha, E. Supriyati, And T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," *Jiko (Jurnal Informatika Dan Komputer)*, Vol. 5, No. 1, Pp. 1-6, Apr. 2022, Doi: 10.33387/Jiko.V5i1.3376.
- [6] M. P. Munthe, A. Siswo, R. Ansori, And R. R. Septiawan, "Analisis Sentimen Komentar Pada Saluran Youtube Food Vlogger Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes," In *E-Proceeding Of Engineering*, 2021.
- [7] M. A. Saddam, E. K. Dewantara, And A. Solichin, "Sentiment Analysis Of Flood Disaster Management In Jakarta On Twitter Using Support Vector Machines," *Sinkron*, Vol. 8, No. 1, Pp. 470-479, Jan. 2023, Doi: 10.33395/Sinkron.V8i1.12063.
- [8] M. Agus Arianto And A. Solichin, "Analisis Sentimen Motogp Mandalika Pada Twitter Menggunakan Metode Naïve Bayes," *Jurnal Ticom: Technology Of Information And Communication*, Vol. 11, No. 1, 2022, [Online]. Available: <https://t.co/Xynw7stiwq>
- [9] R. P. Sidiq, B. A. Dermawan, And Y. Umaidah, "Sentimen Analisis Komentar Toxic Pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," *Jurnal Informatika Universitas Pamulang*, Vol. 5, No. 3, Pp. 356-363, 2020.
- [10] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *Jurnal Media Informatika Budidarma*, Vol. 4, No. 3, Pp. 552-558, 2020, Doi: 10.30865/Mib.V4i3.2140.

- [11] A. Novantirani, M. K. S. S. T, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding of Engineering*, vol. 2, no. 1, pp. 1177–1183, 2015.
- [12] M. R. Adrian, M. P. Putra, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM pada Analisis Sentimen PSBB," *Informatika UPGRIS*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.
- [13] DQLab, "Tahapan Text Preprocessing dalam Teknik Pengolahan Data," *DQLab*, Jun. 21, 2021. <https://www.dqlab.id/tahapan-text-preprocessing-dalam-teknik-pengolahan-data> (accessed Feb. 18, 2023).