

Klasifikasi Sentimen Masyarakat di Twitter Terhadap Ancaman Resesi Ekonomi 2023 dengan Metode Naïve Bayes Classifier

Dea Ropija Sari*, Yusra, Muhammad Fikry, Febi Yanto, Fitri Insani

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950121678@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id, ⁴febiyanto@uin-suska.ac.id, ⁵fitri.insani@uin-suska.ac.id

Email Penulis Korespondensi: 11950121678@students.uin-suska.ac.id

Submitted: 07/06/2023; Accepted: 25/06/2023; Published: 30/06/2023

Abstrak—Resesi ekonomi merupakan suatu kondisi di mana pemulihan ekonomi suatu negara melambat atau memburuk, yang dapat berlangsung selama bertahun-tahun karena pertumbuhan produk domestik bruto (PDB) melambat secara signifikan selama dua kuartal. Peringatan dini akan terjadinya resesi global menjadi perhatian setiap negara di dunia, bahkan resesi global berdampak besar bagi Indonesia. Misalnya, pengeluaran publik dipotong karena pendapatan menurun, pengangguran meningkat dan kemiskinan meningkat, dan banyak yang menghadapi PHK atau pemotongan upah. Penguatan ekonomi penting untuk meminimalisir ancaman tersebut, dan penelitian ini harus dilakukan untuk melihat reaksi masyarakat terhadap ancaman resesi ekonomi. Twitter memberikan wadah kepada pengguna untuk memberikan komentar terhadap permasalahan resesi ekonomi 2023 yang bisa dijadikan informasi klasifikasi sentimen guna mendapatkan komentar positif dan negatif. Penelitian ini menggunakan algoritma naïve bayes classifier. Dalam penelitian ini terdapat tujuh proses utama, yaitu pengumpulan data, pelabelan manual, processing, pembobotan fitur (tf-idf), thresholding, klasifikasi metode naïve bayes, dan pengujian. Dari 1408 data komentar dalam Twitter mengenai ancaman resesi ekonomi 2023. Berdasarkan hasil klasifikasi, menggunakan 2 model pengujian yaitu data balance dan data tidak balance diperoleh hasil pengujian data balance terbaik dengan hasil akurasi tertinggi dengan proses klasifikasi menggunakan algoritma naïve bayes classifier menghasilkan akurasi sebesar 78% yang didapat dengan memakai model perbandingan 90% training data dan 10% test data.

Kata Kunci: Resesi; Klasifikasi Sentimen; Naïve Bayes Classifier; Twitter; Ekonomi

Abstract—Economic recession is a condition in which the economic turnover of a country changes to slow or bad that can last for years as a result of the growth of the Gross Domestic Product (GDP) a country decreases over two decades significantly. Early warnings of the emergence of a global recession become a concern for all countries in the world, even global recessions also have a major impact on Indonesia. Such as declining public spending due to decreasing incomes, increasing unemployment, increasing poverty, and many of whom have to accept PHK or salary cuts. Economic strengthening will be important in minimizing these threats, this research needs to be done to see the response of the public to the threat of economic recession. Twitter provides a container to users to comment on the problem of the economy recession 2023 which can be used as sentiment classification information to know positive and negative comments. This research uses the naïve bayes classifier algorithm. In this study there are seven main processes, namely data collection, manual labelling, processing, feature weighing (tf-idf), thresholding, naïve bayes method classification, testing. From the 1408 comments data on Twitter about the threat of a 2023 economic recession. Based on the results of the classification, using 2 testing models namely data balance and non-balance data obtained the best balance data test results with the highest accuracy result with the process of classification using algoritma naïve bayes classifier resulted in accurateness of 78% obtainable by using a comparison of 90% training data and 10% test data.

Keywords: Recession; Sentiment Classification; Naïve Bayes Classifier; Twitter; Economy

1. PENDAHULUAN

Pada tahun 2020, Indonesia bahkan seluruh dunia terpapar virus Covid-19 yang berdampak pada pertumbuhan ekonomi negara bahkan ekonomi global [1]. Hal ini menyebabkan peningkatan pengangguran dan penurunan pendapatan. Hal ini memicu resesi ekonomi. Resesi adalah fenomena dimana PDB riil negatif selama dua kuartal berturut-turut, yang berarti aktivitas ekonomi, termasuk produksi, distribusi, konsumsi, investasi, dan lain-lain, turun sehingga menimbulkan efek domino yang merugikan berbagai pihak termasuk pemutusan hubungan kerja (PHK)[2]. Ramalan resesi tahun 2023 banyak diberitakan dan menjadi topik perbincangan di media sosial. Mengenai ini jelas saja memicu pro dan kontra di masyarakat.

Pengguna media sosial telah tersebar luas di berbagai lapisan masyarakat dan menyebar dengan sangat cepat. Tidak saja menjadi alat untuk bersosialisasi dan bercengkrama, namun juga untuk memberitahu keinginan dan memvisualisasikan peristiwa dan perasaan masyarakat [3]. diantara jejaring sosial salah satu yang paling populer di masyarakat adalah Twitter. Twitter adalah media sosial yang menghubungkan banyak orang di seluruh dunia dalam berbagai topik [4]. Dengan 328 juta pengguna, Twitter adalah media sosial paling populer di antara berbagai media sosial. Jumlah pengguna Twitter dapat digunakan untuk mengetahui pendapat masyarakat tentang resiko resesi ekonomi tahun 2023. Perihal ini dapat dijadikan sebagai model guna mengetahui pendapat masyarakat tentang dampak masyarakat terhadap resesi ekonomi.

Analisis sentimen secara otomatis memproses data guna mendapatkan sentimen kalimat. Perihal ini dilakukan guna mendeteksi opini atau trend opini publik, apakah lebih ke negatif atau positif. Target dari penelitian ini adalah untuk menilai opini publik terhadap ancaman resesi ekonomi pada tahun 2023. Penelitian ini

menggunakan kata kunci “resesi” dalam tweet berbahasa Indonesia yang diekstraksi dari Twitter dan diproses dengan algoritma klasifikasi naïve bayes classifier untuk membuat klasifikasi ancaman. Resesi ekonomi pada tahun 2023.

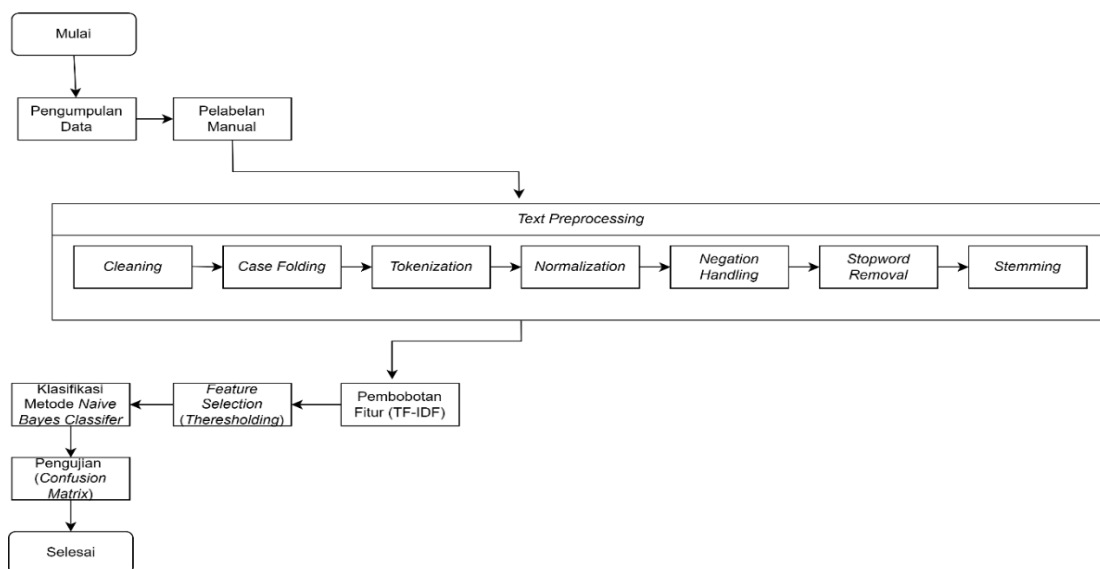
Algoritma klasifikasi naïve bayes classifier dipilih sebab mempunyai kecepatan pemrosesan dan akurasi yang besar apabila memakai data yang cukup besar dan variabel [5]. Metode klasifikasi naïve bayes classifier adalah metode klasifikasi yang digunakan dalam klasifikasi sentimen pada text mining. Metode ini dapat cocok untuk klasifikasi menurut akurasi dan data komputasi. Metode klasifikasi naïve bayes classifier adalah tipe sederhana dari metode naïve bayes classifier yang tepat guna mengklasifikasikan teks atau dokumen [6]. Dalam tugas akhir ini, akan melakukan pengklasifikasian dengan metode klasifikasi naïve bayes classifier. Misi dari penelitian ini ialah untuk mengklasifikasikan secara otomatis opini masyarakat di media sosial Twitter tentang ancaman resesi ekonomi pada tahun 2023 menggunakan data dalam jumlah besar memakai metode naïve bayes classifier.

Penelitian sebelumnya pada tahun 2022 yang diteliti oleh A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, menggunakan metode naïve bayes classifier menganalisis sentimen metode naïve bayes classifier menganalisis sentimen terhadap pembobolan data pada Twitter. metode tersebut memberikan tingkat yang cukup baik dan hasil yang baik. Hasil klasifikasi tertinggi didapatkan dengan nilai accuracy 98.33%, precision 100.00%, dan recall 97.13% [7]. penelitian lainnya pada tahun 2022 yang diteliti oleh Fitriani, E. Utami, and A. Dwi Hartantoand, dengan algoritma naïve bayes classifier dan decision tree dengan permasalahan terkait analisis sentimen terhadap pelaksanaan P3K guru. Hasil akurasi dari kedua algoritma tersebut yaitu naïve bayes classifier memperoleh akurasi sejumlah 100.00% [8]. Penelitian berikutnya pada tahun 2018 yang diteliti oleh B. Gunawan, H. Sasty, P., E. Esyudha, P, juga memakai metode naïve bayes classifier menganalisis sentimen pada ulasan produk. Hasil pengujian tertinggi diperoleh dalam 90% data latih dan 10% data uji dengan nilai akurasi 77.78% [9].

Berdasarkan penelitian selanjutnya pada tahun 2018 yang diteliti oleh B. Mas Pintoko and K. Muslim, memakai metode naïve bayes classifier dengan topik analisis sentimen opini masyarakat terhadap jasa transportasi online. Hasil penelitian ini didapatkan akurasi opini positif sebesar 88.60% dan opini negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil membuktikan tingkat sentimen positif dari tweets lebih tinggi daripada tingkat sentimen negatif [10]. Penelitian terakhir pada tahun 2022 yang diteliti oleh A. M. Anjani, A. A. Chamid, and A. C. Murti, membahas analisis sentimen kaum lgbt pada media sosial Twitter menggunakan algoritma naïve bayes classifier. Berdasarkan penelitian ini memperoleh hasil akurasi sebesar 95% data polaritas positif 77% dan data polaritas negatif sebesar 150% [11].

2. METODOLOGI PENELITIAN

Metode penelitian digambarkan pada diagram fase penelitian guna menyampaikan aturan yang pasti, tersusun dan analitis. Berikut langkah-langkah proyek penelitian, yang menjadi pedoman dalam mengerjakan proses tugas akhir ini. Strategi metodologi penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Flowchart Metodologi Penelitian

Keterangan:

Tahapan pertama yang dilakukan yaitu pengumpulan data, setelah data terkumpul selanjutnya melakukan pelabelan data melalui validasi seorang lulusan sarjana bahasa Indonesia, setelah semua data diberi label tahapan

ketiga melakukan preprocessing, tahapan selanjutnya yaitu pembobotan kata, tahapan kelima Feature Selection, setelah melalui Feature Selection maka tahapan selanjutnya yaitu klasifikasi metode naïve bayes classifier, dan yang terakhir melakukan pengujian menggunakan (confusion matrix).

2.1 Pengumpulan Data

Pengumpulan data adalah langkah pertama dalam metodologi penelitian. Data yang dikumpulkan berasal dari sosial media Twitter dengan keyword ‘resesi’. Data yang didapatkan berjumlah 1408 data. Setelah mengumpulkan data tahapan selanjutnya melakukan pelabelan manual atau kerap disebut dengan labeling di proses guna mendefinisikan kategori sentimen positif atau negatif [12]. Setelah dilakukan pelabelan manual kemudian divalidasi oleh validator seorang sarjana lulusan bahasa Indonesia bapak Robby Fitrianto, Spd.

2.2 Preprocessing

Teks preprocessing merupakan langkah umum dalam klasifikasi [13], Preprocessing memiliki peran yang sangat penting dalam sistem Natural Language Processing (NLP) karena tahap ini bertanggung jawab untuk mengidentifikasi karakter, kata, dan kalimat yang menjadi bagian dasar yang akan diolah pada seluruh bagian pemrosesan selanjutnya, mulai dari analisis morfologi hingga penandaan ucapan.[14]. Selanjutnya masuk ke bagian evaluasi dan uji. Ditahap ini data yang total awalnya 1 data. Dengan tujuan menghasilkan term/kata dari data yang cocok dengan kebutuhan [15]. Teks preprocessing pada penelitian ini, merupakan sekelompok bagian secara berurutan yaitu cleaning, case folding, Tokenization, normalization, negative handling, stopword removal, stemming. sebagaimana penguraian dari bagian-bagian tersebut sebagai berikut :

1. Cleaning merupakan proses menghapus karakter yang tidak perlu seperti angka, hastag, symbol, mention, dan link url dari proses klasifikasi.
2. Case Folding merupakan mengganti seluruh huruf membentuk huruf kecil sehingga dihitung seperti kata yang serupa [16].
3. Tokenization merupakan dalam tokenisasi, kalimat dibagi menjadi unit leksikal yang lebih kecil yang disebut kata. Tujuan dari tokenisasi adalah memeriksa kata-kata dalam sebuah kalimat [17].
4. Normalization merupakan bagian untuk mengganti kata-kata yang tidak baku menjadi bentuk baku. Hal tersebut dibuat guna menurunkan jumlah fitur yang dihitung per kata [18].
5. Negation Handling merupakan proses ketika terdapat kata bersifat negasi yaitu tidak, bukan, nnggak, dan lainnya hal ini akan menyebabkan arti kata menjadi berlawanan dengan arti sebenarnya.
6. Stopword Removal adalah membuang kata-kata yang memiliki kuantitas kemunculan yang sering tapi tidak terlalu berpengaruh. Kata yang masuk stopword yaitu “dan”, “di”, “yang”, “dari” sehingga hanya meninggalkan kata-kata yang berharga [19].
7. Stemming berguna untuk menjadikan kata sebagai kata alami atau kata dasar, dengan membuang semua imbuhan yang ada dalam kata tersebut [20].

2.3 Pembobotan Kata

Term Frequency-Inverse Document Frequency (TF-IDF) adalah sebuah metode algoritma yang menambah bobot dalam teks. Tf merupakan berapa kali kata tersebut terlihat pada dokumen, sementara IDF merupakan nilai kebalikan dari dokumen yang menyimpan kata tersebut. Ketika TF dan IDF dikalikan akan mendapatkan nilai bobot pada kata tersebut [21]. di bawah ini adalah rumus persamaan TF-IDF:

$$W_{t,d} = T_{Ft,d} \cdot I_{DFt} \tag{1}$$

$$I_{DFt} = \log (N / n_t) \tag{2}$$

$$TF - IDF_{t,d} = T_{Ft,d} \cdot I_{DFt} \tag{3}$$

Keterangan :

$T_{Ft,d}$ = kerapatan kehadiran t (term) pada dokumen d

$W_{t,d}$ = bobot terhadap t (term) dalam dokumen

$TF - IDF_{t,d}$ = pembobotan kalimat (d) terhadap kata (t)

I_{DFt} = kebalikan frekuensi dokumen

N = jumlah dokumen n_t = jumlah dokumen mengandung kata t

2.4 Klasifikasi Naïve Bayes Classifier

Naïve bayes classifier dipakai guna menggali nilai probabilitas terbaik pada kategori yang paling tepat dalam teknik klasifikasi data uji [22]. Selama proses klasifikasi, metode klasifikasi naïve bayes classifier menggali probabilitas terbaik di antara bagian dokumen yang diuji, dibawah ini menunjukkan perhitungannya.

$$P(C|X) = \frac{P(C) \prod_{i=1}^n P(X_i|C)}{P(X)} \tag{4}$$

Keterangan :

X= menguji data dari kategori yang tidak terlihat

P(X) = probabilitas X

P(C) = probabilitas C

X = kelas yang terdefinisi dengan baik

C = kumpulan data

P(X|C) = probabilitas X untuk dengan situasi hipotesis C

P(C|X) = probabilitas hipotesis C karena keadaan

2.5 Pengujian

Pengujian pada tugas akhir ini mempunyai misi untuk menemukan performa atau kinerja dari tipe yang dianjurkan. Metode yang dipakai pada pengujian dalam tugas akhir ini adalah confusion matrix. Confusion matrix adalah tabel yang memberikan laporan komparatif antara hasil klasifikasi yang sesungguhnya dengan hasil diproses oleh sistem (prediksi). Dalam confusion matrix, terdapat 4 sebutan nilai hasil klasifikasi yang selanjutnya akan dipakai guna menghitung nilai akurasi, recall, presisi, dan fl-score, yaitu true positive (TP), true negative (TN) adalah hasil kasifikasi yang benar. Nilai false positive (FP) merupakan nilai dimana hasilnya diprediksi positif tetapi ternyata negatif sementara false negative (FN) merupakan nilai dimana hasil prediksinya negatif tetapi ternyata positif [23].

Tabel 1. Pengujian

	Perkiraan Positif	Perkiraan Negatif
Nyata Positif	TP	FN
Nyata Negatif	FP	TN

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \tag{5}$$

$$Presisi = \frac{TP}{TP+FP} * 100\% \tag{6}$$

$$Recall = \frac{TP}{TP+FN} * 100\% \tag{7}$$

$$F - Measure = \frac{Precision . Recall}{Precision+Recall} \tag{8}$$

Keterangan:

TP = True Positive

FN = False Negative

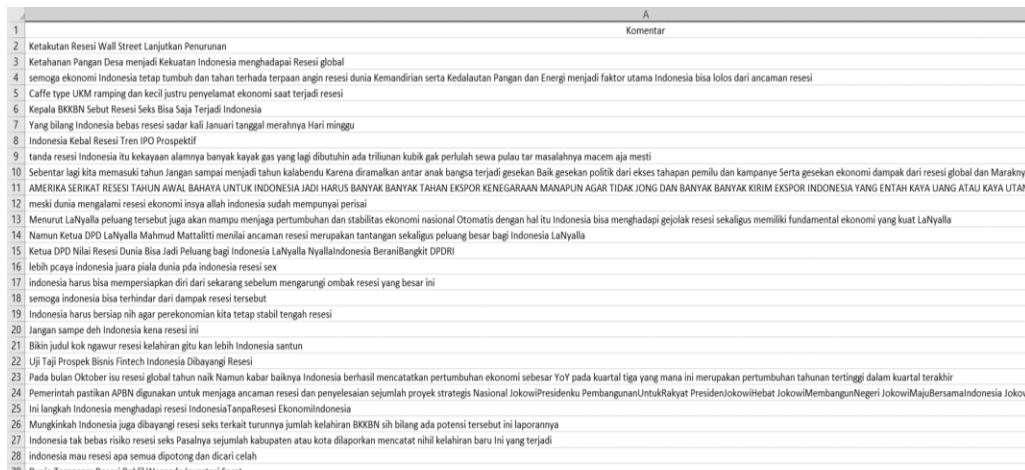
FP = False Positive

TN = True Negative

3. HASIL DAN PEMBAHASAN

3.1 Data Set

Data set yang dipakai dalam tugas akhir ini merupakan dataset yang berasal dari jejaring sosial yaitu Twitter berupa tweets yang berisi kata “resesi” data yang diambil dari bulan Oktober-Desember 2022. Pada penelitian ini, terdapat 1408 data tweets yang diklasifikasikan kedalam 2 kategori ialah negatif dan positif dutujkan pada gambar 2 sebagai berikut.



	A
1	Komentar
2	Ketakutan Resesi Wall Street Lanjutkan Penurunan
3	Ketahanan Pangan Desa menjadi Kekuatan Indonesia menghadapi Resesi global
4	semoga ekonomi Indonesia tetap tumbuh dan tahan terhadap terpaan angin resesi dunia Kemandirian serta Kedaulatan Pangan dan Energi menjadi faktor utama Indonesia bisa lolos dari ancaman resesi
5	Caffe type UKM ramping dan kecil justru penyelamat ekonomi saat terjadi resesi
6	Kepala BKKBN Sebut Resesi Seks Bisa Saja Terjadi Indonesia
7	Yang bilang Indonesia bebas resesi sadar kali Januari tanggal merahnya Hari minggu
8	Indonesia Kegal Resesi Tren IPO Prospektif
9	tanda resesi Indonesia itu kekayaan alamnya banyak kayak gas yang lagi dibutuhkan ada triliunan kubik gak perlulah sewa pulau tar masalahnya macem aja mesti
10	Sementara lagi kita memasuki tahun Jangan sampai menjadi tahun kalabendu Karena diramalkan antar anak bangsa terjadi gesekan Baik gesekan politik dari eksekutif dan kampanye Serta gesekan ekonomi dampak dari resesi global dan Maraknya
11	AMERIKA SERIKAT RESESI TAHUN AWAL BAHAYA UNTUK INDONESIA JADI HARUS BANYAK BANYAK TAHAN EKSPOR KENEGARAAN MANAPUN AGAR TIDAK JONG DAN BANYAK BANYAK KIRIM EKSPOR INDONESIA YANG ENTAH KAYA UANG ATAU KAYA UTAN
12	meski dunia mengalami resesi ekonomi insya allah Indonesia sudah mempunyai perisai
13	Menurut LaNyalla peluang tersebut juga akan mampu menjaga pertumbuhan dan stabilitas ekonomi nasional Otomatis dengan hal itu Indonesia bisa menghadapi gejala resesi sekaligus memiliki fundamental ekonomi yang kuat LaNyalla
14	Namun Ketua DPD LaNyalla Mahmud Mattalitti menilai ancaman resesi merupakan tantangan sekaligus peluang besar bagi Indonesia LaNyalla
15	Ketua DPD Nilai Resesi Dunia Bisa Jadi Peluang bagi Indonesia LaNyalla NyallaIndonesia BeraniBangkit DPDR
16	lebih percaya Indonesia juara piala dunia pda indonesia resesi sex
17	Indonesia harus bisa mempersiapkan diri dari sekarang sebelum mengaruhi ombak resesi yang besar ini
18	semoga Indonesia bisa terhindar dari dampak resesi tersebut
19	Indonesia harus bersiap nih agar perekonomian kita tetap stabil tengah resesi
20	Jangan sampe deh Indonesia kena resesi ini
21	Bikin judul kok ngawur resesi kelahiran gitu kan lebih Indonesia santun
22	Uji Taji Prospek Bisnis Fintech Indonesia Dibayangi Resesi
23	Pada bulan Oktober isu resesi global tahun naik Namun kabar baiknya Indonesia berhasil mencatatkan pertumbuhan ekonomi sebesar YoY pada kuartal tiga yang mana ini merupakan pertumbuhan tahunan tertinggi dalam kuartal terakhir
24	Pemerintah pastikan APBN digunakan untuk menjaga ancaman resesi dan penyelesaian sejumlah proyek strategis Nasional JokowiPresidenku PembangunanUntukRakyat PresidenJokowiHebat JokowiMembangunNegeri JokowiMajuBersamaIndonesia Jokowi
25	Ini langkah Indonesia menghadapi resesi IndonesiaTanpaResesi EkonomiIndonesia
26	Mungkinkah Indonesia juga dibayangi resesi seks terkait turunnya jumlah kelahiran BKKBN sih bilang ada potensi tersebut ini laporannya
27	Indonesia tak bebas risiko resesi seks Pasalnya sejumlah kabupaten atau kota dilaporkan mencatat nihil kelahiran baru ini yang terjadi
28	Indonesia mau resesi apa semua dipotong dan dicari celah
29	Dunia Terancam Resesi Bahili Wasoada Investasi Seret

Gambar 2. Dataset

3.1 Preprocessing Data

Setelah mengumpulkan data sesuai dengan kebutuhan maka akan dilakukan preprocessing teks untuk mengubah data mentah menjadi data proses. Ada beberapa langkah dalam tahap preprocessing data dapat dilihat pada tabel 2 sebagai berikut.

Tabel 2. Hasil Preprocessing

Nama	Input	Output
Cleaning	Sudah banyak pembahasan tentang resesi tapi Indonesia sendiri optimis bisa menghadapi dgn faktor beberapa yang Indonesia punya	Sudah banyak pembahasan tentang resesi tapi Indonesia sendiri optimis bisa menghadapi dgn faktor beberapa yang Indonesia punya
Case Folding	Sudah banyak pembahasan tentang resesi tapi Indonesia sendiri optimis bisa menghadapi dgn faktor beberapa yang Indonesia punya	sudah banyak pembahasan tentang resesi tapi indonesia sendiri optimis bisa menghadapi dgn faktor beberapa yang indonesia punya
Tokenization	sudah banyak pembahasan tentang resesi tapi indonesia sendiri optimis bisa menghadapi dgn faktor beberapa yang indonesia punya	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dgn, faktor, beberapa, yang, indonesia, punya,
Normalization	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dgn, faktor, beberapa, yang, indonesia, punya,	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, yang, indonesia, punya,
Negation Handling	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, yang, indonesia, punya,	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, yang, indonesia, punya,
Stopword Removal	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, yang, indonesia, punya,	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, indonesia, punya,
Stemming	sudah, banyak, pembahasan, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, menghadapi, dengan, faktor, beberapa, Indonesia, punya	sudah, banyak, bahas, tentang, resesi, tapi, indonesia, sendiri, optimis, bisa, hadap, dengan, faktor, berapa, indonesia, punya,

Setelah proses preprocessing maka akan dilakukan visualisasi dengan menggunakan word cloud. word cloud adalah gambaran visual dari seberapa sering sebuah kata muncul pada satu set dokumen. Ukuran huruf menunjukkan seberapa sering suatu kata muncul, jadi semakin besar ukuran hurufnya maka semakin sering kata itu muncul. Kebalikannya, semakin kecil ukuran huruf, semakin berkurang kata tersebut muncul [24]. Word cloud dalam tugas akhir ini ditunjukkan pada gambar 3 yaitu Word cloud negatif dan gambar 4 Word cloud positif.



Gambar 3. Word Cloud Negatif



Gambar 4. Word Cloud Positif

3.2 Pembobotan Kata

Selanjutnya ke tahap pembobotan kata, setiap kata atau term yang sudah melalui preprocessing akan diberi bobot melalui TF-IDF. Berikut tampilan hasil perhitungan TF-IDF menggunakan tools google collab ditunjukkan pada gambar 5 sebagai berikut.

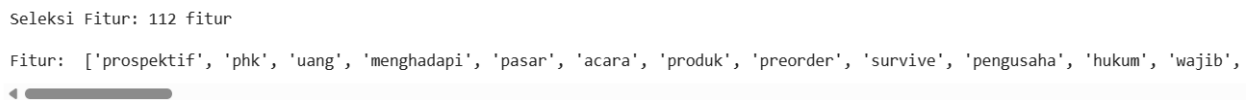
0s [23]	aamiin	aamin	abadi	abai	abang	abdi	abdul	abdulhayat	abdullah	\
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
2667	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2668	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2669	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2670	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2671	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

0	abdulmuthalib	...	zenix	zero	zinah	ziyech	zona	zonauang	zoom	\
0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
2667	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2668	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2669	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2670	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2671	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 5. Hasil TF-IDF

3.3 Feature Selection

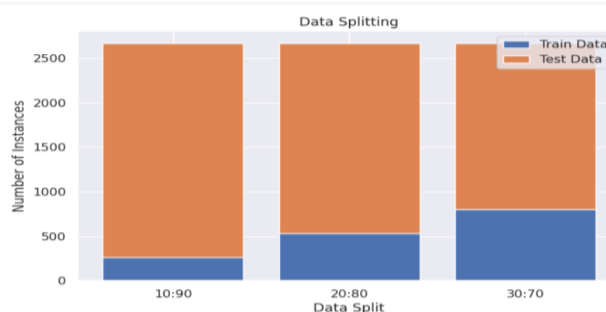
Selanjutnya ke tahap pemilihan fitur terbaik menggunakan thresholding. Pada Penelitian ini memakai ambang batas yaitu 0,001. Berikut gambar 6 tampilan hasil thresholding.



Gambar 6. Hasil Thresholding

3.4 Klasifikasi Naïve Bayes Classifier

Sebelumnya data telah melewati preprocessing, pelabelan dan (TF-IDF) menjadi bagian dalam proses training dan testing. Langkah pertama yang harus dilakukan adalah memisahkan training data dan test data. Seperti ditunjukkan dalam gambar 7 berikut.



Gambar 7. Train Data dan Test Data

Langkah selanjutnya melakukan klasifikasi menggunakan metode naïve bayes classifier untuk membuat acuan yang dapat digunakan untuk memproses klasifikasi data tanpa pelabelan manual. Pengujian ini diproses memakai metode confusion matrix yang akan dijadikan kedalam perhitungan accuracy, recall, dan precision. Berikut gambaran dari hasil klasifikasi metode naïve bayes classifier pada gambar 8 memakai tools Google Collab.

Klasifikasi Naive Bayes Classifier																				
i	abang	abdi	abis	acara	action	acu	ada	adadikompa	...	zaman	zero	zinah	ziyech	zona	zonauang	zulhas	zulkifli	Sentimen	Hasil Klasifikasi	
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Positif	Positif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Positif	Negatif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Negatif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Positif	Positif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Negatif
...
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Negatif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Positif	Negatif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Negatif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Positif
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Negatif	Positif

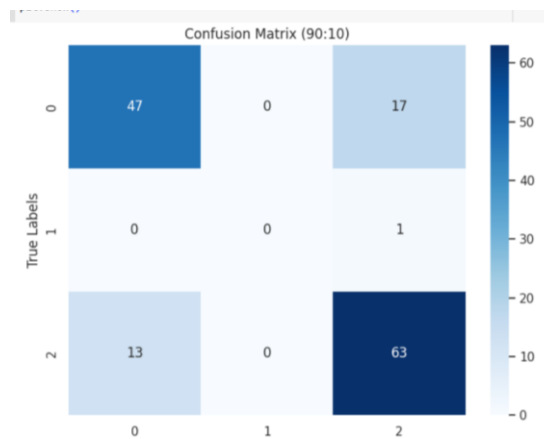
Gambar 8. Hasil Klasifikasi Metode Naïve Bayes Classifier

3.5 Pengujian

Pada penelitian ini melakukan pengujian untuk menguji train data dan test data dari 1408 data dengan menggunakan confusion matrix. Penelitian ini juga menggunakan dua pengujian yaitu data balance dengan 704 data positif dan 704 data negatif, yang kedua yaitu data tidak balance dengan 464 data positif dan 704 data negatif selanjutnya melakukan tiga perbandingan yang pertama dimana training data dan test data 90:10, perbandingan selanjutnya training data dan test data 20:80, dan perbandingan yang terakhir training data dan test data 70:30.

3.6.1 Pengujian Data Balance

Pengujian ini menggunakan 1408 data yang diklasifikasikan data positif berjumlah 704 dan data negatif berjumlah 704. dibawah ini merupakan hasil pengujian menggunakan confusion matrix.



Gambar 9. Hasil Confusion Matrix Data Balance Perbandingan 90:10

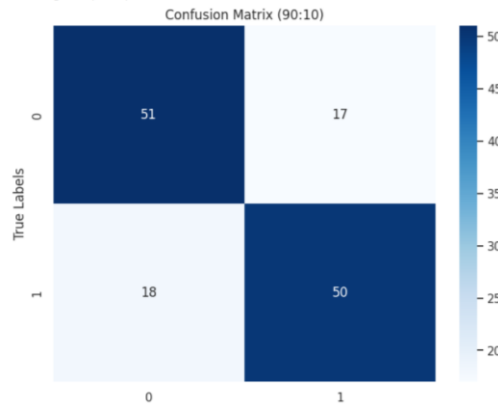
Hasil confusion matrix percobaan training data dan test data 90:10, jadi memperoleh nilai akurasi sebanyak :

$$\begin{aligned}
 \text{Akurasi} &= \frac{47 + 63}{47 + 63 + 17 + 13} * 100\% \\
 &= \frac{110}{140} = 0,78 * 100\% = 78\%
 \end{aligned}$$

Metode naïve bayes classifier memperoleh hasil akurasi dengan perbandingan training data dan test data 90:10 sebanyak 78%. Menurut pengujian yang telah diproses hasil akurasi terbaik diperoleh di perbandingan training data dan test data 90:10 dengan accuracy 78%, presisi 77%, recall 78%, dan f1-score 78%.

3.6.2 Pengujian Data Tidak Balance

Pengujian ini menggunakan 1353 data yang diklasifikasikan data positif berjumlah 646 dan data negatif berjumlah 704. dibawah ini hasil pengujian memakai confusion matrix.



Gambar 10. Hasil Confusion Matrix Data Tidak Balance Perbandingan 90:10

Hasil confusion matrix percobaan training data dan testing data 90:10, jadi memperoleh nilai akurasi sebanyak :

$$\begin{aligned}
 \text{Akurasi} &= \frac{51 + 50}{51 + 50 + 17 + 18} * 100\% \\
 &= \frac{101}{136} = 0,74 * 100\% = 74\%
 \end{aligned}$$

Metode naïve bayes classifier memperoleh hasil akurasi dengan perbandingan training data dan test data 90:10 sebanyak 78%. Menurut pengujian yang telah diproses hasil akurasi terbaik diperoleh di perbandingan training data dan test data 90:10 dengan accuracy 74%, presisi 74%, recall 74%, dan f1-score 74%.

4. KESIMPULAN

Berdasarkan hasil penerapan dan pengujian klasifikasi sentimen masyarakat di Twitter terhadap ancaman resesi ekonomi tahun 2023 memakai metode naïve bayes classifier dari jumlah 1408 data, maka dapat diambil kesimpulan. Dengan model pengujian data balance perbandingan seluruh train data dan test data mendapatkan accuracy tertinggi pada perbandingan 90% train data dan 10% test data sehingga pengujian data balance ini mendapatkan hasil tertinggi accuracy 78%, presisi 77%, recall 78%, dan f1-score 78%. yang menyatakan metode naïve bayes classifier cukup akurat untuk bisa mengatasi klasifikasi sentimen masyarakat di Twitter terhadap ancaman resesi ekonomi 2023. Namun Ketika datanya dikurangi dan klasifikasi datanya tidak balance maka akurasi juga akan rendah. Saran untuk penelitian setelah ini, bisa memakai metode klasifikasi lain untuk melihat hasil dari kinerja metode tersebut menggunakan permasalahan yang sama yaitu klasifikasi sentimen masyarakat terhadap ancaman resesi 2023. Kemudian bisa memakai data mentah dari jejaring sosial lainnya yaitu Youtube, Instagram, atau Facebook.

REFERENCES

- [1] S. Ardila, S. Annatasha, V. Aulya D. Kustiawati, J. Pendidikan and D. Konseling, “Studi Literatur : Mempersiapkan Investasi untuk Hadapi Isu Resesi Ekonomi di Indonesia,”2022.
- [2] S. Blandina, A. Noor Fitriani, and W. Septiyani, “Strategi Menghindarkan Indonesia dari Ancaman Resesi Ekonomi di Masa Pandemi,” Efektor, vol. 7, no. 2, pp. 181–190, Dec. 2020, doi: 10.29407/e.v7i2.15043.
- [3] A. N. Made Ayu Juli, D. G. Hendra Divayana, and G. Indrawan, “Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier,” Jurnal Sistem dan Informatika (JSI), vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.
- [4] A. Wibisono, S. Dadi Rizkiono, and A. Wantoro, “Filtering Spam Email Menggunakan Metode Naive Bayes,” Telefortech, (Journal Of Telematics and Information Technology), vol. 1, no.1,pp 9-17 Nov. 2020.
- [5] L. Aji Andika and P. Amalia Nur Azizah, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” Indonesian Journal of Applied Statistics, vol. 2, no. 1, pp May 2019.
- [6] W. A. Prabowo and C. Wiguna, “Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM,” Jurnal Media Informatika Budidarma, vol. 5, no. 1, p. 149, Jan. 2021, doi: 10.30865/mib.v5i1.2604.
- [7] A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, “Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes,” Jurnal Teknologi Informatika dan Komputer, vol. 8, no. 2, pp. 202–213, Sep. 2022, doi: 10.37012/jtik.v8i2.1240.
- [8] Fitriani, E. Utami, and A. Dwi Hartantoand, “Analisis Sentimen Masyarakat Terhadap Pelaksanaan P3K Guru dengan Algoritma Naïve Bayes Classifier Dan Decision Tree“ Teknologi informasi dan multimedia, vol. 3, no. 1, Mei 2022.
- [9] B. Gunawan, H. Sasty, P., E. Esyudha, P., “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com

- [10] B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier." e-Proceeding of Engineering, vol. 5, no.3 pp Desember 2018.
- [11] A. M. Anjani, A. A. Chamid, and A. C. Murti, "Analisis Sentimen Kaum Lgbt Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," Jurnal Teknik Informatika vol. 1, no. 2, pp. Juli, Halaman 1~8, 2022.
- [12] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 2, pp. 231–237, Apr. 2021, doi: 10.29207/resti.v5i2.2698.
- [13] I. Kurnia Syuriadi and W. Astuti, "Klasifikasi Teks Multi Label pada Hadis dalam Terjemahan Bahasa Indonesia Berdasarkan Anjuran, Larangan dan Informasi menggunakan TF-IDF dan KNN," e-Proceeding of Engineering, vol. 06, no. 02, Agustus 2019.
- [14] M. Alif, I. Aulia, and Y. E. Kurniawati, "Pengembangan Model Klasifikasi Dokumen Artikel Teks Berita Olahraga dan Bukan Olahraga dalam Bahasa Indonesia Menggunakan Algoritma Support Vector Machine," Jurnal Mahasiswa Institut Teknologi dan Kalbis vol. 8, no. 2, Mei 2022.
- [15] E. T. Handayani and A. Sulistiyawati, "Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes," Jurnal Teknologi dan Sistem Informasi (JTSI), vol. 2, no. 3, pp. 32–37, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [16] L. A. Mullen, K. Benoit, O. Keyes, D. Selivanov, and J. Arnold, "Fast, Consistent Tokenization of Natural Language Text," J Open Source Softw, vol. 3, no. 23, p. 655, Mar. 2018, doi: 10.21105/joss.00655.
- [17] R. Ghaniaviyanto Nur, "Indonesian Online News Topics Classification using Word2Vec and K-Nearest Neighbor," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 6, pp. 1083–1089, Dec. 2021, doi: 10.29207/resti.v5i6.3547.
- [18] A. Rafid Rizqullah, A. Wedhasmara, R. Izwan Heroza, A. Putra, and P. Putra, "Analisis Masalah Pada Data Review Aplikasi Terhadap Layanan E-Commerce Menggunakan Metode Text Classification." Jurnal TEKNO KOMPAK, vol. 16, no. 1, 2022.
- [19] A. Deka and M. Sharma, "A Comparative Analysis of Vegetation Radiometric Indices for Classification of Bambusa Tulda using Satellite Imagery," International Journal of Computer Sciences and Engineering Open Access, vol. 7, no. 1, Desember 2019.
- [20] B. Kaida Palma, D. Triantoro Murdiansyah, and W. Astuti, "Klasifikasi Teks Artikel Berita Hoaks Covid-19 dengan Menggunakan Algotrima K-Nearest Neighbor." E-proceeding of Engineering, vol. 8, no. 5, Oktober 2021.
- [21] B. K. Hananto, A. Pinandito, and A. P. Kharisma, "Penerapan Maximum TF-IDF Normalization Terhadap Metode KNN Untuk Klasifikasi Dataset Multiclass Panichella Pada Review Aplikasi Mobile," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer vol. 2, no 12, Desember 2018.
- [22] Y. A. Rahman, "Vaksinasi Massal Covid-19 sebagai Sebuah Upaya Masyarakat dalam Melaksanakan Kepatuhan Hukum (Obedience Law)," Khazanah Hukum, vol. 3, no. 2, pp. 80–86, Apr. 2021, doi: 10.15575/kh.v3i2.11520.
- [23] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter" Jurnal Stiki Informatika, vol. 10, no. 02, Desember 2020.
- [24] R. Parluka, S. Ilham Pradika, A. M. Hakim, and R. N. M. Kholilul, "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob," Jurnal Ilmiah Teknologi Informasi dan Robotika, vol 02, no. 02, Desember 2020.