

Analisis Sentimen Masyarakat Terhadap Resesi Ekonomi Global 2023 Menggunakan Algoritma Naïve Bayes Classifier

Sriani¹, Aidil Halim Lubis², Yunus Fadillah Harahap³

^{1,2,3} Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Indonesia

e-mail: sriani@uinsu.ac.id¹, aidilhalimlubis@uinsu.ac.id², yunusharahap5@gmail.com³

ARTICLE INFO

Article history:

Received 18 Maret 2022

Received in revised form 20 Maret 2022

Accepted 28 Maret 2022

Available online 2 Juli 2022

ABSTRACT

The global economic recession is a global economic downturn that affects the domestic economies of countries in the world. The stronger the economic dependence of one country on the global economy, the faster a recession will occur in that country. In 2020 the country of Indonesia and even the world are exposed to the COVID-19 virus which has an impact on the country's economic growth, even the world economy. This is the trigger for an economic recession. This has led to many different public perspectives on the occurrence of a global economic recession whose opinions or reactions are expressed on social media Youtube. The data was obtained by crawling techniques from social media Youtube with a total of 500 comments used. The data is then labeled (class) with a lexicon-based method with an Indonesian language dictionary. From the labeling results, it was obtained 185 positive labeled data (37%) and 315 negative opinions (63%). The data preprocessing stage is carried out in preparation for the data to be processed for sentiment analysis. Of the many opinions obtained, an analysis of public sentiment regarding the 2023 global economic recession will be carried out using the Naïve Bayes classification algorithm. This study also applied the TF-IDF word weighting method with the n-gram feature used, namely bigram (n=1). The system will be evaluated using a confusion matrix. The implementation results show a prediction model with a total of 500 opinion data with a comparison of training data and test data of 9:1, producing an accuracy value of 84.00%, a precision value of 75.00%, a recall of 30.00%, and an f1-score of 42.86%. The performance of the system model built in this study can be said to be good.

Keywords: Global Economic Recession; Sentiment Analysis; Naïve Bayes Classifier

1. Pendahuluan

Resesi merujuk pada sebuah perihal tentang ekonomi negara mengalami penurunan, dapat dilihat berdasarkan angka Produk Domestik Bruto (PDB) yang minus, peningkatan tingkat tunakarya, serta perkembangan ekonomi yang sesungguhnya mengalami pemerosotan selama dua kuartal berturut-turut atau dalam satu tahun. Penyebab utama terjadinya resesi dalam ekonomi global pada tahun 2023 dapat dikaitkan dengan adanya krisis keuangan di beberapa negara, yang dipicu oleh ketidakstabilan dalam pasar keuangan. ^[1]

Dikutip dari CBNC Indonesia.com pada Jumat (30/09/2022), Presiden Joko Widodo (Jokowi) memberikan peringatan mengenai situasi dunia yang berada dalam kondisi sulit dan dihadapkan pada tantangan besar pada tahun mendatang. Beliau mencermati adanya 'awan gelap' dan badai besar yang akan dihadapi. Presiden Jokowi menekankan kehati-hatian terhadap tingkat ketidakpastian, sering kali diingatkan setiap harinya melalui berbagai media, baik sosial, cetak, maupun online. Isu-isu seputar resesi global dan kesulitan ekonomi tahun ini, menurutnya, menjadi pokok pembicaraan, dan beliau kembali menyampaikan bahwa tahun depan diprediksi akan lebih sulit, dengan ketidakpastian besar mengenai dimensi sejati dari badai yang akan datang.^[2] Dikutip dari laman berita tempo.co pada Minggu (08/01/2023), Menteri Keuangan Sri Mulyani mengonfirmasi bahwasannya Indonesia terkait resesi dapat dipastikan tidak mungkin mengalaminya di tahun 2023. Pernyataannya didasarkan proyeksi ekonomi global yang baru-baru ini diumumkan oleh Dana Moneter Internasional (IMF). "IMF merilis dugaan bahwa kemungkinan perekonomian dunia mengalami resesi. Alhamdulillah, kita tidak termasuk dalam hal tersebut, bahkan harus menjaga stabilitas ekonomi," ujarnya.^[3]

Analisis sentimen merupakan domain yang melibatkan pemrosesan bahasa alami, linguistik komputasional, dan penambangan teks. Fokusnya adalah mengevaluasi komentar, aksi, pengiraan, sentimen, serta evaluasi terhadap narasumber yang terkait atas berbagai entitas seperti produk, topik, organisasi, individu, layanan, atau aktivitas. Analisis sentimen bermanfaat untuk menilai apakah tanggapan pengguna terhadap suatu produk positif atau negatif dengan cara mengekstrak informasi emosional dari ulasan. Naïve Bayes Classifier digunakan dalam pendekatan untuk analisis sentimen, sebuah struktur klasifikasi probabilistik simpel dalam perhitungan probabilitas melalui perhitungan frekuensi serta gabungan nilai dari sebuah dataset. Keunggulan metode ini terletak pada kemampuannya untuk melakukan pengklasifikasian dengan memanfaatkan besaran terhadap data training termasuk relatif kecil demi memastikan tujuan pengukuran proses klasifikasi yang dibutuhkan^[5].

Berdasarkan penjelasan di atas, rencananya akan dilakukan studi untuk menganalisis sentimen masyarakat terkait dengan resesi ekonomi global pada tahun 2023 menggunakan metode klasifikasi algoritma Naïve Bayes. Algoritma Naïve Bayes Classifier digunakan pada penelitian ini guna mengelompokkan opini dari komentar-komentar, bertujuan supaya dapat memisahkan opini kelas positif atau negatif. Hasil dari penelitian ini diharapkan akan membentuk sebuah model klasifikasi supaya dapat menganalisis sentimen dan memberikan informasi mengenai sentimen yang terdapat dalam komentar-komentar tersebut.

2. Metode Penelitian

Metode pengumpulan data yang dilakukan melibatkan komentar pengguna YouTube yang terdapat pada konten berjudul "2023, Ekonomi Dunia Gelap Gulita?" di saluran CNN Indonesia. Data ini diperoleh melalui teknik web crawling dan pemanfaatan Youtube API. Proses web crawling dilaksanakan dengan menggunakan alat Google Colaboratory dan memanfaatkan pustaka bahasa pemrograman Python, yaitu pandas. Data yang dikumpulkan mencakup komentar pengguna, tanggal unggahan komentar, dan nama pengguna YouTube, dan kemudian disimpan dalam format .csv. Meskipun total data yang diperoleh melebihi 3000, hanya 500 data komentar yang akan diolah dalam studi.

Pre-processing data merupakan langkah yang penting pada pemrosesan data mining, sebab data atau atribut tidak semua dapat digunakan. Tujuan dari proses ini adalah untuk menyelaraskan data yang akan digunakan agar sesuai dengan kebutuhan analisis. Berikut hasil tahap - tahapan proses preprocessing pada penelitian ini, yaitu :

1. Cleaning

Dalam tahapan ini, semua komentar akan mengalami pembersihan dengan upaya membuang sebuah komponen tidak ada memberikan nilai, seperti nama pengguna, angka, hashtag, tanda baca, URL atau tautan, dengan tujuan pengurangan noise yang ada dalam suatu data. ^[8].

2. Case Folding

Dalam tahap ini, seluruh huruf pada dataset diseragamkan menjadi huruf kecil (lowercase). ^[9].

3. Tokenizing

Langkah ini melibatkan pengecekan dari karakter pertama hingga karakter terakhir. Apabila karakter pertama tidak termasuk karakter pemisah kata, seperti titik (.), koma (,), spasi, atau tanda pemisah lainnya, dengan demikian karakter tersebut akan disatukan atau digabungkan dengan karakter berikutnya.

4. Normalization

Normalization adalah tahapan dimana memperbaiki, menghapus kata yang tidak baku, singkatan, serta bentuk penulisan kata yang tidak sesuai dengan kerangka ataupun kata yang tepat dalam Bahasa Indonesia.

5. Stopword Removal

Dalam tahap ini, setiap kata dalam komentar akan diperiksa, dan kemudian dilakukan penghapusan kata yang diduga kurang relevan . Misalnya, kata sambung, kata depan, kata ganti, atau kata yang tidak berhubungan dengan analisis sentimen akan dihilangkan ^[10].

6. Stemming

Dalam langkah ini, data tweet akan mengalami modifikasi dengan menghilangkan imbuhan, sehingga setiap kata akan diwakili sebagai kalimat dasar atau bentuk dasarnya. ^[8].

Tabel 1. Hasil Tahap *Preprocessing*

<i>Cleaning</i>	Stop smtr bangun IKN agar Indonesia lolos dr resesi ekonomi
<i>Case folding</i>	stop smtr bangun ikn agar indonesia lolos dr resesi ekonomi
<i>Tokenizing</i>	'stop', 'smtr', 'bangun', 'ikn', 'agar', 'indonesia', 'lolos', 'dr', 'resesi', 'ekonomi'
<i>Normalization</i>	berhenti sementara bangun ikn supaya indonesia lolos dari resesi ekonomi
<i>Stopword Removal</i>	'berhenti', 'bangun', 'ikn', 'indonesia', 'lolos', 'resesi', 'ekonomi'
<i>Stemming</i>	henti bangun ikn indonesia lolos resesi ekonomi

3.3 Labelisasi

Tahapan ini dimana akan dilakukan pemberian label pada *dataset* yang dibagi dalam dua kelas. Untuk komentar yang memiliki sentimen positif akan diberikan label positif, sebaliknya komentar yang memiliki sentimen negatif akan diberikan label negatif. Sehingga dapat diberikan label berdasarkan nilai yang terdapat masing – masing komentar. Dalam proses labelisasi sentimen yang dilakukan secara otomatis dengan menggunakan *lexicon-based*.

Pendekatan berbasis leksikon juga merupakan metode umum untuk analisis sentimen di media sosial karena bersifat praktis. Metode ini menggunakan kamus sebagai sumber bahasa atau leksikon. Pendekatan berbasis leksikon bertindak sebagai Pengklasifikasi sentimen dari setiap pendapat memfasilitasi pengelompokan kalimat sentimen ke dalam kategori negatif dan positif. Dengan penerapan metode ini, proses klasifikasi data kalimat sentimen dapat dilakukan dengan lebih mudah.^[11]

Setelah setiap kata dalam teks diberi label, maka dapat dihitung skor sentimen keseluruhan teks dengan menghitung jumlah kata positif dan negatif, dan menggabungkan nilai ini secara matematis. Formula yang umum digunakan untuk menghitung skor sentimen (*StSc*) adalah^[12]:

$$StSc = \frac{\text{jumlah kata positif} - \text{jumlah kata negatif}}{\text{total jumlah kata}} \dots\dots\dots(1)$$

3.4 Pembobotan TF-IDF

Dalam langkah ini, dilakukan perhitungan bobot untuk pada suatu kata (term) yang terdapat dalam suatu dataset yang berasal dari komentar pengguna YouTube. Proses TF-IDF dimulai dengan menghitung nilai term frequency (TF) dan document frequency (DF). Selanjutnya, dihitung nilai invers document frequency (IDF) dan bobot akhir diperoleh dengan perkalian nilai TF dan IDF. Formulasnya yaitu :

$$IDF = \ln \frac{d+1}{df+1} + 1 \dots\dots\dots(2)$$

$$W_{dt} = TF_{dt} \times IDF \dots\dots\dots(3)$$

Setelah itu , nilai *Wdt* dinormalisasikan untuk menyamakan bobotnya sehingga berada dalam rentang yang seragam. Formula untuk normalisasi nilai Norm *Wdt* dapat dijelaskan melalui persamaan (4) berikut :

$$Norm W_{dt} = \frac{W_{dt}}{\sqrt{\sum_{i=1}^n x_i^2}} \dots\dots\dots(4)$$

3.5 Split Dataset

Setelah melalui tahapan pre-processing dan ekstraksi fitur dengan TF-IDF, langkah selanjutnya adalah melakukan Split dataset. Proses ini melibatkan pemilihan dan pembagian dataset menjadi dua, yakni data latih dan data uji. Dalam skenario ini, data dipisahkan dengan proporsi dimana jumlah data latih lebih besar daripada data pengujian. Dalam penelitian ini, data splitting dengan rasio 9:1, dimana 90% dari setiap dataset untuk data latih, sedangkan 10% sisanya untuk data uji. Prinsip ini juga berlaku untuk skenario data splitting lainnya. Tujuan dari pembagian ini adalah untuk menilai kinerja model dengan menggunakan data uji yang tidak terlibat dalam proses pelatihan.

3.6 Naïve Bayes Classifier\

Naïve Bayes digunakan perhitungan probabilitas dan menerapkan Teorema Bayes, yang dirumuskan dengan formula berikut ini:

$$P(T|X) = \frac{P(X|T)P(T)}{P(X)} \dots\dots\dots(5)$$

Langkah-langkah dalam menghitung klasifikasi kelas pada data uji dokumen pertama dimulai dengan menghitung nilai *prior probability*, *conditional probability*, dan selanjutnya *posterior probability* :

- 1. Menghitung nilai *prior probability*

$$P(X) = \frac{dx}{d} \dots\dots\dots(6)$$

Dimana, P(X) adalah *Prior probability*, x adalah Hipotesis data adalah suatu kategori (kelas), dengan d menyatakan jumlah total dataset dan dx menyatakan jumlah data yang termasuk dalam kategori (kelas) x.

- 2. Menghitung nilai *conditional probability* dengan *Laplacian smoothing*

$$P(Term H_n|X) = \frac{H_n|X + 1}{(c) + |V|} \dots\dots\dots(7)$$

Dimana, P(Term $H_n|X$) adalah *Conditional probability* kata ke-n dengan kelas x, ($H_n|X$) adalah kemunculan kata dalam suatu kelas x diukur melalui frekuensinya, (c) adalah Bobot (*IDF*) dari kata pada kelas x, dan $|V|$ adalah Nilai *IDF* dari seluruh kata yang terdapat dalam dokumen tersebut.

- 3. Menghitung nilai *posterior probability*

$$P(Term H|X) = P(X) * P(Term H_1|X) * P(Term H_2|X) \dots\dots\dots(8)$$

Dimana, P(Term H|X) adalah *posterior probability*, P(X) adalah *prior probability*, P(Term $H_n|X$) adalah *conditional probability*

3.7 Evaluasi Sistem

Pada tahapan dalam penelitian ini, Confusion matrix digunakan untuk melakukan perhitungan tingkat akurasi pada dataset dengan menyajikan hasil klasifikasi dalam bentuk tabel. *Confusion matrix* yang diperoleh berupa jumlah data uji yang telah berhasil diprediksi. Hasilnya berbentuk matriks berordo 2x2 sesuai dengan jumlah kelas sentimen positif dan negatif. Hasil *confusion matrix* menghitung nilai *accuracy*, *recall*, dan *precision*, dan *f1-score* dari sistem yang telah dibangun dalam proses klasifikasi menggunakan metode. Berikut tampilan tabel hasil *confusion matrix* pada penelitian ini :

Tabel 2. Hasil Klasifikasi *Confusion Matrix*

		Kelas Sebenarnya	
		Positif	Negatif
Kelas Prediksi	Positif	3	7
	Negatif	1	39

Dari Tabel 2 maka besaran nilai *accuracy*, *precision*, *recall*, dan *f1-score* bisa terhitung :

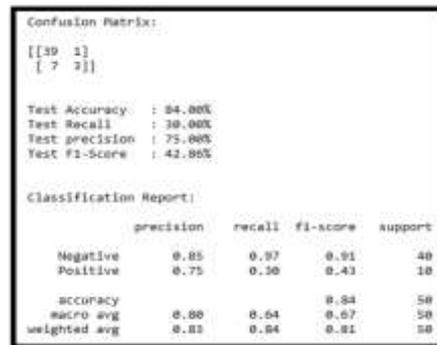
$$\text{Accuracy} = \frac{3 + 39}{3 + 1 + 7 + 39} * 100\% = 84.00 \%$$

$$\text{Precision} = \frac{3}{3 + 7} * 100\% = 30.00 \%$$

$$\text{Recall} = \frac{3}{3 + 7} * 100\% = 75.00 \%$$

$$f1 - score = \frac{2 \times 75 \times 30}{75 + 30} * 100\% = 42.8571429 \% = 42.86 \%$$

Classification report memperlihatkan hasil dari seluruh nilai yang telah disajikan sebelumnya. Selanjutnya hasil *classification report* dari *confusion matrix* :



```

Confusion Matrix:
[[39  1]
 [ 7  1]]

Test Accuracy : 84.00%
Test Recall   : 30.00%
Test Precision: 75.00%
Test F1-Score : 42.85%

Classification Report:

```

	precision	recall	f1-score	support
Negative	0.85	0.97	0.91	40
Positive	0.75	0.30	0.43	10
accuracy			0.84	50
macro avg	0.80	0.64	0.67	50
weighted avg	0.83	0.64	0.81	50

Gambar 2. *Classification Report*

4. Kesimpulan

Dengan merujuk pada hasil penelitian analisis sentimen masyarakat terhadap resesi ekonomi global pada tahun 2023 menggunakan Naïve Bayes Classifier, dapat disimpulkan bahwa dataset yang diperoleh dengan teknik crawling didapat sejumlah +3000 data komentar(opini) mengenai resesi ekonomi global 2023. Data yang digunakan dalam menerapkan preprocessing data, labelisasi sebanyak 500 data komentar(opini) yang akan dianalisis. Dalam penelitian ini, data dilabeli menggunakan Teknik berbasis lexicon dengan kamus berbahasa Indonesia. Pendekatan ini mengukur sejauh mana kata-kata yang terdeteksi dalam kamus sesuai dengan dokumen (opini) dan menghitung skornya. Dengan penerapan teknik ini, rincian data pada kategori positif mencapai 37%, yaitu terdapat 185 data opini yang termasuk dalam kategori tersebut, sementara kategori negatif mencapai 63%, atau 315 data opini. Metode Naïve Bayes Classifier Tingkat menghasilkan tingkat akurasi dapat dianggap baik pada sistem klasifikasi sentimen. Hal ini terlihat dari hasil akurasi pada dataset yang terdiri dari 500 dataset dengan selisih data training dan data testing sebesar 9:1, terdiri dari 450 data training dan 50 data testing. Dengan menggunakan skema pembobotan kata yang disebut TF-IDF, diperoleh nilai akurasi sebesar 84,00%, presisi sebesar 75,00%, recall bernilai 30,00%, dan f1-score bernilai 42,86%.

Penelitian ini diharapkan dapat menjadi dasar atau acuan bagi penelitian yang akan dilakukan di masa mendatang. Dapat melakukan analisis sentimen terhadap data yang mengandung emotikon, gambar, maupun kalimat sarkasme. Percobaan dengan menggunakan metode pembobotan kata lainnya Dengan tujuan untuk menentukan metode yang lebih efektif dalam analisis sentimen, penelitian ini juga dapat membuka jalan bagi pengembangan sistem yang berkaitan dengan perbaikan antarmuka pengguna. Antarmuka ini dapat berbasis web,

mobile, atau desktop, bertujuan untuk mempermudah pengguna lain dalam membaca hasil analisis sentimen.

References

- [1] Blandina, S. R., Fitriani, A. N., & Septiyani, W. (2020). Strategi Menghindarkan Indonesia dari Ancaman Resesi Ekonomi di Masa Pandemi. *Efektor*, 7(2), 181–190. <https://doi.org/10.29407/e.v7i2.15043>
- [2] Redaksi, T. Merinding! Ini Pidato Jokowi, Sri Mulyani & Luhut Soal Resesi. CNBN Indonesia [Internet]. [cited 2023 Mei 11]. Available form: <https://www.cnbcindonesia.com/news/20220930065738-4-376136/merinding-ini-pidato-jokowi-sri-mulyani-luhut-soal-resesi>
- [3] Putri, R. S. Sri Mulyani Pastikan 2023 Indonesia Tidak Terkena Resesi: Insya Allah Kita Jaga Terus. Tempo.Co [Internet]. [cited 2023 Mei 11]. Available form: <https://bisnis.tempo.co/read/1677026/sri-mulyani-pastikan-2023-indonesia-tidak-terkena-resesi-insya-allah-kita-jaga-terus>
- [4] Agustina, N. C. ., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54.
- [5] Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika*, 6(1), 18–22. <https://doi.org/10.31311/ji.v6i1.4685>
- [6] Toepak, E. W., Arwani, I., & Afirianto, T. (2018). Pembangunan Aplikasi Penyedia Informasi Lowongan Pekerjaan Menggunakan Youtube API Pada Smartphone Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 168–174.
- [7] Anggraini, M., Tyas, R. A., Sulasiyah, I. A., & Aini, Q. (2020). Implementasi Algoritma Naïve Bayes Dalam Penentuan Rating Buku. *Sistemasi*, 9(3), 557. <https://doi.org/10.32520/stmsi.v9i3.915>
- [8] Fajri, A. R. H., Roy, B. S., Mubarak, H., Pangestu, A. D., & Prasvita, D. S. (2021). Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Naïve Bayesian Classification dan Support Vector Machine. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA) Jakarta-Indonesia, September*, 166–175.
- [9] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50–59. <https://doi.org/10.35314/isi.v3i1.335>
- [10] Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Analisis Sentimen Masyarakat Terhadap Resesi Ekonomi Global 2023 Menggunakan Algoritma Naïve Bayes Classifier (Sriani)

Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.

- [11] Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. *Komputek*, 3(2), 52. <https://doi.org/10.24269/jkt.v3i2.270>
- [12] Aline, B. *Lexicon-Based Sentiment Analysis: A Tutorial*. Knime. [Internet]. [cited 2023 Mei 11]. Available form: <https://doi.org/https://www.knime.com/blog/lexicon-based-sentimen-analysis>