

**Perbandingan Metode *Support Vector Machine* dan *Naïve Bayes* Pada Analisis Sentimen Terhadap Isu Resesi 2023
Pada Media Sosial Twitter Dan Komentar Youtube**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Faiz Aunullah
NIM:402019611010



TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS DARUSSALAM GONTOR
PONOROGO
2023

PENGESAHAN

Penerapan Metode *Support Vector Machine* dan *Naïve Bayes* Pada Analisis Sentimen Terhadap Isu Resesi 2023 Pada Media Sosial Twitter Dan Komentar Youtube

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
Memperoleh gelar Sarjana Komputer

Disusun Oleh :
Faiz Aunullah
NIM : 402019611010

Skripsi ini telah diuji dan dinyatakan lulus pada
16 April 2023
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I

Dosen Pembimbing II

Aziz Musthafa M.T
NIY: 150487

Jumhurul Umami M.Sc.
NIY: 060198

Dosen Penguji I

Dosen Penguji II

Dihin Muriyatmoko, M.T
NIY: 150489

Oddy Virgantara Putra, M.T
NIY: 160589

Mengetahui
Ketua Prodi Teknik Informatika

Dihin Muriyatmoko, M.T.
NIY: 150489

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Ponorogo, 16 April 2023

Faiz Aunullah

NIM: 402019611010

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yaitu Allah SWT, karena dengan rahmat dan karunia-Nya kita dapat merasakan kegembiraan dan kesempatan belajar. Sholawat dan salam selalu ditujukan kepada Nabi Muhammad SAW, sebagai utusan-Nya yang telah menunjukkan jalan yang lurus dan terang seperti yang kita rasakan saat ini.

Dan tak lupa kami ucapkan terima kasih yang sebesar-besarnya kepada semua dosen yang telah membimbing kami selama proses penulisan skripsi ini. Dengan segala kerendahan hati, penulis menyadari bahwa dalam proses pengerjaan skripsi ini selalu ada dan tidak dapat dipisahkan peran dari pihak-pihak yang telah membantu, mengarahkan dan mendorong. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ustadz Profesor. Dr. Hamid Fahmy Zarkasyi, M.A, M.Phil., selaku Rektor beserta seluruh jajarannya, serta seluruh dosen Universitas Darussalam Gontor Ponorogo yang telah mendidik penulis.
2. Ustadz Haris Setyaningrum, M.Sc, selaku Dekan Fakultas Sains dan Teknologi serta para dosen.
3. Ustadz Dihin Muriyatmoko, M.T., selaku Ketua Program Studi Teknik Komputer beserta dosen dan stafnya, yang mendorong dan membimbing penulis untuk melakukan penelitian untuk skripsi ini.
4. Ustadz Aziz Musthafa S.Kom, M.T, dan Ustadz Jumhurul Umami M.sc, selaku pembimbing yang telah bersedia meluangkan waktunya untuk membimbing, mendukung, dan mengarahkan Penulis.
5. Ustadz Muhammad Taqiyuddin, S.H.I., M.Ag, selaku dosen dan pembimbing Islamisasi Ilmu, yang bersedia meluangkan waktunya untuk berkontribusi dan mengintegrasikan ilmu keislaman ke dalam penelitian.
6. Terima kasih yang sebesar-besarnya kepada kedua orang tua dan keluarga besar yang selalu mendukung, mendoakan dan memberi semangat tanpa henti.
7. Teman dari program sarjana Teknik Informatika masih mendukung dan membantu penelitian ini.

Penulis menemukan bahwa masih banyak kekurangan dalam penelitian ini. Oleh karena itu, tim penulis sangat mengharapkan kritik dan saran untuk menjadikan penelitian ini lebih lengkap dan lengkap, bermanfaat bagi perkembangan ilmu pengetahuan khususnya di bidang teknologi informasi sektor publik. Semoga tugas akhir ini dapat membawa manfaat bagi penulis pada khususnya dan pada umumnya kepada semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT senantiasa memberikan rahmat dan hidayah-Nya kepada kita, Aamiin ya rabbal 'alamin

Ponorogo, 16 April 2023

Faiz Aunullah

faizaunullah@unida.gontor.ac.id

ABSTRAK

Resesi adalah sebuah istilah yang digunakan untuk menggambarkan keadaan dimana perputaran ekonomi suatu negara berubah menjadi lambat atau buruk. Perputaran ekonomi yang melambat ini bisa berlangsung cukup lama bahkan tahunan akibat dari pertumbuhan produk domestik bruto (PDB) suatu negara menurun selama dua kuartal dan berlangsung secara terus menerus. Adanya resesi sangat merugikan perekonomian yang ada pada suatu daerah apalagi pada tahun 2023 terdengar isu akan terjadinya resesi di negara Indonesia. Terdapat berbagai macam opini masyarakat yang di ungkapkan di media sosial tentang isu resesi 2023 pada data twitter dan juga komentar pada youtube. Isu tersebut menimbulkan pro dan kontra lebih lagi resesi tersebut sangat merugikan perekonomian Indonesia. Pada penelitian ini memiliki tujuan untuk mengklasifikasikan sentimen masyarakat yang di ungkapkan pada media sosial twitter dan youtube tentang isu resesi yang terjadi pada tahun 2023 dengan kumpulan tweet dan komentar pada youtube dalam opini positif, negatif dan netral menggunakan metode *Naïve Bayes*, dan *Support Vector Machine*, Model yang dibuat menggunakan metode *Naïve Bayes* menghasilkan nilai akurasi sebesar 0.724, dan menggunakan metode *Support Vector Machine* menghasilkan nilai akurasi sebesar 0.800. Pada penelitian ini metode *Support Vector Machine*(SVM) Memiliki tingkat akurasi yang lebih baik dibandingkan metode *Naïve Bayes* Diharapkan untuk peneliti selanjutnya menambahkan lebih banyak jumlah data, karena dataset akan sangat mempengaruhi hasil dari model dan akurasi pada klasifikasi. Memodifikasi pada tahapan *pre-processing* sehingga dapat meningkatkan hasil akurasi.

Kata kunci: Resesi, Analisis Sentimen, *Confussion Matrix*

ABSTRACT

Recession is a term used to describe a situation in which the economic turnover of a country changes to slow or worse. This slowing economic turnover can last long enough even yearly as a result of a country's gross domestic product (GDP) growth declining for two quarters and continuously continuing. There is a recession that is very detrimental to the economy that exists in a region, especially in 2023 it seems that the issue of recession will occur in Indonesia. There are a variety of public opinions expressed on social media about the issue of the recession 2023 on Twitter data and also comments on YouTube. This issue raises the pros and cons, and the recession has greatly damaged the Indonesian economy. The study aims to classify the public sentiment expressed on social media twitter and youtube about the recession issue that occurred in 2023 with the collection of tweets and comments on youtube in positive, negative and neutral opinions using the Naïve Bayes method, and Support Vector Machine. A model made using the Naïve Bayes method yields an accuracy value of 0.724, and using the Support Vector Machine method produces an accurate value of 0,800. In this study, the Support Vector Machine (SVM) method has a better level of accuracy than the Naïve Bayes method. It is expected for future researchers to add more amounts of data, as the datasets will greatly affect the outcome of the model and the accuracy on the classification. Modify at the pre-processing stage so that the accuracy results can be improved.

Keywords: Resession, Emotional Analysis, Confusion Matrix

DAFTAR ISI

PENGESAHAN	i
PERNYATAAN ORISINALITAS	ii
KATA PENGANTAR	iii
ABSTRAK.....	v
<i>ABSTRACT</i>	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
BAB 1 PENDAHULUAN	1
1.1 Latar belakang.....	1
1.2 Rumusan masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Batasan masalah.....	4
1.6 Sistematika pembahasan	5
BAB 2 KAJIAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.1.1 Ezza Ramadhanta Machmud Razaq, Deden Witarsyah Jacob, Faqih Hamami(2021)	6
2.1.2 Andreansyah, Grace Gata(2022)	7
2.1.3 Cholid Fadilah Hasri, Debby Alita(2022)	7
2.1.4 Debby Alita, RB Ali Shodiqin(2023)	8
2.1.5 Dadang Iskandar Mulyana, Nesti Lutfianti(2023).....	9
2.2 Kajian Teori	16
2.2.1 Resesi.....	16
2.2.2 Media Sosial	16
2.2.3 Analisis Sentimen	16
2.2.4 Netlytic	17
2.2.5 Naïve Bayes.....	17
2.2.6 Support Vector Machine.....	18

2.2.7	Confusion Matrix.....	19
BAB 3	METODOLOGI PENELITIAN.....	20
3.1	Waktu dan Tempat Penelitian.....	20
3.1.1	Waktu.....	20
3.1.2	Tempat.....	20
3.2	Analisis Kebutuhan.....	20
3.2.1	Perangkat Keras :.....	21
3.2.2	Perangkat Lunak :.....	21
3.3	Perancangan Sistem.....	21
3.3.1	Pengumpulan dan Pembangunan Dataset.....	23
3.3.2	Pre-Processing.....	23
3.3.3	Pembobotan TF-IDF.....	25
3.3.4	Processing.....	26
3.3.5	Uji Coba dan Evaluasi.....	27
BAB 4	HASIL DAN PEMBAHASAN.....	30
4.1	Hasil.....	30
4.1.1.	Pengumpulan Data dan Pembangunan Dataset.....	30
4.1.2.	Pre-Processing.....	31
4.1.3.	Pembobotan TF-IDF.....	36
4.2	Hasil Klasifikasi Naïve Bayes.....	37
4.3	Hasil Klasifikasi Support Vector Machine.....	38
4.4	Hasil Perbandingan.....	40
4.5	Implementasi Sistem.....	40
4.5.1	Hasil Analisis Penyebaran Tren Data.....	41
BAB 5	Penutup.....	43
5.1	Kesimpulan.....	43
5.2	Saran.....	43
	DAFTAR PUSTAKA.....	44

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	11
Tabel 2.2 <i>Confusion Matrix</i>	19
Tabel 3.1 Timeline Penelitian	20
Tabel 3.2 <i>K-fold Validation</i>	28
Tabel 4.1 Jumlah Data Per Kategori Kelas	30
Tabel 4.2 Contoh Tahapan <i>Case Folding</i>	31
Tabel 4.3 Contoh Hasil <i>Tokenizing</i>	32
Tabel 4.4 Contoh Hasil <i>Filtering</i>	33
Tabel 4.5 Hasil <i>Stemming</i>	34
Tabel 4.6 <i>Precision, Recall Dan F1-Score Naïve Bayes</i>	37
Tabel 4.7 Rincian <i>K- Fold Naïve Bayes</i>	38
Tabel 4.8 <i>Precision, Recall Dan F1-Score SVM</i>	39
Tabel 4.9 Rincian <i>K- Fold Support Vector Machine</i>	39
Tabel 4.10 Hasil Perbandingan	40
Tabel 4.11 Jumlah Data Per Kata Kunci	40
Tabel 4.12 Tren Sentimen Data Per Kata Kunci	42

DAFTAR GAMBAR

Gambar 3.1 Tahapan Perancangan Sistem.....	22
Gambar 3.2 Contoh Undersampling dan Oversampling.....	25
Gambar 3.3 <i>Flowchart</i> Klasifikasi.....	26
Gambar 3.4 3x3 <i>Confusion Matrix</i>	27
Gambar 4.1 Sebelum <i>Random Oversampling</i>	35
Gambar 4.2 Setelah <i>Random Oversampling</i>	36
Gambar 4.3 Naïve Bayes Confussion Matrix	37
Gambar 4.4 <i>SVM Confusion Matrix</i>	38
Gambar 4.5 Tren Sentimen Data Per Tanggal	41
Gambar 4.6 Tren Sentimen Data Per Hari	41

BAB 1 PENDAHULUAN

1.1 Latar belakang

Resesi ekonomi global diperkirakan akan datang pada tahun 2023, dikutip dari cnbc Indonesia, menteri keuangan ibu Sri Mulyani Indrawati SE msc Ph D menyampaikan bahwasanya hampir semua negara akan mengalami resiko kemunduran ekonomi.¹ ekonomi yang gelap di tahun 2023 tak lepas dari ancaman tensi geopolitik antara Rusia dan Ukraina karena negara tersebut termasuk sebagai pemasok pangan dan energi membuat tingkat inflasi di banyak negara meningkat. Resiko resesi disebabkan oleh kenaikan *cost of fund* dan potensi default di banyak negara yang memiliki rasio utang yang sangat tinggi. Harga komoditas yang tinggi kemudian menyebabkan inflasi.²

Saat ini banyak negara yang memperbincangkan akan terjadinya resesi di negara masing masing. Salah satunya adalah Indonesia. Fenomena terjadinya resesi dan depresi terlihat dari melemahnya aktivitas ekonomi sektor riil pada negara tersebut. Resesi dan depresi adalah fenomena biasa dalam suatu perekonomian yang terjadi dalam jangka waktu Panjang yang perekonomiannya terus tumbuh dan merupakan fenomena konjungtur jangka Panjang yang pasti terjadi.³

Setiap individu masyarakat memiliki tanggapan dan opini berbeda beda. Opini seseorang terhadap sesuatu sangatlah penting dan juga merupakan salah satu hal yang mempengaruhi perilaku seseorang dalam bermasyarakat. Sentimen analysis merupakan bidang studi yang menganalisa opini masyarakat, sentimen, evaluasi, penilaian, sikap dan emosi terhadap sebuah pelayanan publik, seorang tokoh dan isu atau masalah serta peristiwa yang terjadi pada masyarakat itu sendiri. Melihat dari hal tersebut Allah SWT telah menerangkan dalam suatu ayat tentang kebesaran nya atas apa yang telah diciptakan, yaitu :

¹ “Sri Mulyani Bilang AS Lolos Dari Resesi 2023, RI Gimana?,” accessed April 18, 2023, <https://www.cnbcindonesia.com/news/20230303100106-4-418558/sri-mulyani-bilang-as-lolos-dari-resesi-2023-ri-gimana>.

² BAMBANG PRIJAMBODO, “Resesi Dunia 2023, Seberapa Parah?,” [www.kompas.id](https://www.kompas.id/baca/opini/2023/02/19/resesi-dunia-2023-seberapa-parah), 2023, <https://www.kompas.id/baca/opini/2023/02/19/resesi-dunia-2023-seberapa-parah>.

³ Bachtiar Hassan Miraza, “Seputar Resesi Dan Depresi,” *Jurnal Ekonomi Kiat* 30, no. 2 (2019): 11–13, <https://journal.uir.ac.id/index.php/kiat>.

وَمَا كَانَ النَّاسُ إِلَّا أُمَّةً وَاحِدَةً فَاخْتَلَفُوا وَلَوْلَا كَلِمَةٌ سَبَقَتْ مِنْ رَبِّكَ لَفُضِيَ بَيْنَهُمْ فِي مَا فِيهِ يَخْتَلِفُونَ

Artinya “Dan manusia itu dahulunya hanyalah satu umat, kemudian mereka berselisih. Kalau tidak karena suatu ketetapan yang telah ada dari Tuhanmu, pastilah telah diberi keputusan (di dunia) di antara mereka, tentang apa yang mereka perselisihkan itu..”(QS Yunus Ayat 19)

Menurut tafsir Al-Mukhtashar / Markaz Tafsir Riyadh di jelaskan bahwasanya tidaklah umat manusia ini dahulu melaikan satu umat yang beriman dan bertauhid. Kemudian mereka berselisih paham. Sebagian dari mereka tetap beriman dan sebagian lainnya berubah menjadi kafir. Sekiranya tidak ada ketentuan yang telah Allah tetapkan sebelumnya bahwa Dia tidak akan memutus perselisihan mereka di dunia, sehingga dapat diketahui dengan jelas siapa yang berada di jalan yang benar dan siapa yang tersesat.

Dalam penelitian ini membahas tentang analisis sentimen masyarakat terhadap isu resesi 2023. Berdasarkan isu yang ramai dibicarakan di media sosial yang mana masalah tersebut sangat merugikan perekonomian Indonesia sebagaimana yang telah terjadi pada tahun 1998 dan juga pada tahun 2008. Banyak pengguna media sosial yang memberikan pendapat terhadap isu resesi 2023. Pendapat tersebut bisa berupa positif, negatif ataupun netral. Masalah ini menarik untuk diteliti guna mengetahui opini masyarakat terhadap isu ini. Pada penelitian ini menggunakan metode *Naïve Bayes* dan *Support Vector Machine*. yang akan digunakan untuk mengklasifikasi data dari twitter dan juga data dari youtube yang diklasifikasikan menjadi kelas positif, negatif, dan netral.

Penggunaan *Machine Learning* disini karena *Machine Learning* dapat membuat sistem otomatis. Dengan algoritma yang terbentuk didalamnya bisa berkerja dengan sesuai perintah yang sudah di tentukan. Dengan hal ini akan mempercepat analisis data yang kita lakukan. Selain itu, *Machine Learning* juga memudahkan sistem oprasional dalam melakukan inovasi produk dan memiliki analisis yang akurat karena *Machine Learning* dapat bekerja dengan mempelajari dan mengoleksi data secara sistematis.

1.2 Rumusan masalah

Rumusan masalah yang ada dalam penelitian ini adalah terdapat berbagai macam opini masyarakat yang diungkapkan di media sosial tentang isu resesi 2023 pada data twitter dan juga komentar pada youtube. Isu tersebut menimbulkan pro dan kontra terlebih lagi resesi tersebut sangat merugikan perekonomian Indonesia.

1.3 Tujuan

Tujuan dari penelitian ini adalah mengklasifikasikan sentimen masyarakat yang diungkapkan pada media sosial twitter dan youtube tentang isu resesi yang terjadi pada tahun 2023 kumpulan tweet dan komentar pada youtube dalam opini positif, negatif dan netral menggunakan metode *Naïve Bayes dan Support Vector Machine*

1.4 Manfaat

Manfaat yang diharapkan penulis dari adanya penelitian ini adalah sebagai berikut :

a) Bagi Kajian Teori

1. Mampu kontribusi positif dalam pemikiran akan dunia sosial yang penuh akan ketidak jelasan
2. Mampu menyimpulkan pandangan masyarakat terhadap resesi 2023 dari kumpulan tweets dalam opini negatif, positif dan netral

b) Bagi Pembaca

1. Peneliti dapat berkontribusi dalam pengembangan akademik maupun non akademik yang terkait akan pendapat masyarakat resesi 2023
2. Menyadarkan masyarakat bahwa adanya resesi pada tahun 2023

c) Bagi Penulis

1. Dapat merancang sistem yang dapat mengklasifikasikan kumpulan tweet dalam opini Negatif, Positif, Netral
2. Mengaplikasikan dan mengamalkan pelajaran *data mining* dan *text mining*, karena *prototype* ini berbasis *Data Scientist* dan machine learning.

d) Bagi Universitas

1. Mengetahui kemampuan mahasiswa dalam penguasaan materi pembelajaran yang diperoleh ketika masa perkuliahan
2. Mengetahui kemampuan mahasiswa dalam mengimplementasikan ilmunya sebagai bahan evaluasi bagi mahasiswa lainnya

1.5 Batasan masalah

Agar permasalahan yang dirumuskan dapat lebih fokus, maka penelitian memiliki Batasan masalah sebagai berikut :

1. Sistem ini masih berupa *prototype*.
2. Bahasa pemrograman menggunakan python.
3. Sistem hanya menggunakan data dari youtube dan twitter dengan kata kunci resesi 2023 yang diambil secara acak dari tanggal 10 Oktober 2022 sampai 3 Januari 2023
4. Data yang diambil sebanyak 5504 data dengan pembagian 2004 dari twitter dan 3500 dari data komentar youtube
5. Sistem hanya dapat digunakan untuk klasifikasi pandangan masyarakat terhadap isu resesi 2023 yang banyak dibicarakan di media sosial twitter dan komentar pada youtube yang di bagi menjadi kelas positif, negatif, dan netral
6. Proses klasifikasi menggunakan metode Naïve Bayes, dan Support Vector Machine.
7. Proses Pengumpulan dan pelabelan data tidak membedakan personal yang berkomentar pada twitter dan youtube

1.6 Sistematika pembahasan

Sistematika penulisan pada penelitian ini terdiri dari lima bagian yaitu :

BAB I : Pendahuluan

- 1.1 Latar Belakang
- 1.2 Rumusan Masalah
- 1.3 Tujuan
- 1.4 Batasan Masalah
- 1.5 Manfaat
- 1.6 Sistematika Pembahasan

BAB II : Kajian Pustaka

- 2.1 Penelitian Terdahulu
- 2.2 Kajian Teori

BAB III : Metode Penelitian

- 3.1 Waktu dan Tempat Penelitian
- 3.2 Alat dan Bahan Penelitian
- 3.3 Tahapan Penelitian dan Metode

BAB IV : Hasil dan Pembahasan

- 4.1 Latar Belakang
- 4.2 Rumusan Masalah

BAB V : Penutup

- 1.1 Kesimpulan
- 1.2 Saran

BAB 2 KAJIAN PUSTAKA.

2.1 Penelitian Terdahulu

2.1.1 Ezza Ramadhanta Machmud Razaq, Deden Witarsyah Jacob, Faqih Hamami(2021)⁴

Penelitian ini dilakukan pada oktober 2021 dengan judul “Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid 19 Pada Media Sosial Twitter menggunakan Perbandingan Algoritma Klasifikasi” dalam penelitian ini dengan latar belakang pandemi covid-19 yang menyebabkan pembatasan aktivitas manusia yang mempengaruhi segala aspek salah satunya dalam bidang Pendidikan. Kegiatan mengajar berubah dari offline menjadi online yang ditetapkan oleh pemerintah. sehingga muncul beberapa opini dan pandangan yang diungkapkan (mahasiswa) melalui twitter oleh sebab itu peneliti melakukan analisis sentimen untuk mengetahui kecenderungan komentar para pelajar terhadap regulasi pembelajaran di Indonesia apakah positif, negatif, atau netral. Untuk mengetahui tanggapan pengguna twitter terhadap pembelajaran online selama pandemi, peneliti menggunakan algoritma klasifikasi yang memiliki tingkat akurasi tinggi, penulis akan membandingkan 3 algoritma klasifikasi yaitu, algoritma klasifikasi *Naïve Bayes*, *K-Nearest Neighbor (K- NN)* dan *C4.5*. Adapun tahapan yang dilakukan pada penelitian ini adalah dengan cara preprocessing data, processing data, klasifikasi, dan evaluasi.

Setelah membandingkan ketiga algoritma tersebut didapatkan akurasi tertinggi yaitu 60.80% menggunakan algoritma klasifikasi K-Nearest Neighbor (K-NN). Hasil akurasi yang didapat menggunakan ekstraksi fitur TF-IDF sebesar 60.80%, presisi (specificity) 40.73%, recall (sensitivity) 72%, dan F1-Measure 38.43%. Analisis sentimen yang didapat pada penelitian ini menunjukkan bahwa pengguna twitter di Indonesia lebih banyak memberikan komentar netral.

⁴ Ezza Ramadhanta Machmud Razaq, Jacob Deden Witarsyah, and Faqih Hamami, “Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi,” *E-Proceeding of Engineering* 8, no. 5 (2021): 9000–9006.

2.1.2 Andreansyah, Grace Gata(2022)⁵

Penelitian ini dilakukan pada tahun 2022 dengan judul “Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN” Media sosial sebagai tempat untuk mengakses dan menyebarkan informasi telah berkembang sangat pesat, salah satunya adalah twitter. Twitter pada penelitian ini digunakan sebagai sumber untuk mengetahui pendapat atau opini masyarakat terhadap kebijakan pemerintah tentang vaksin booster sebagai syarat mudik di Indonesia tahun 2022. Opini-opini tersebut bisa berupa opini positif atau negatif, tergantung dari pandangan publik terhadap objek tersebut. Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen dengan menggunakan algoritma K-Nearest-Neighbors, untuk mengetahui seberapa besar akurasi yang dihasilkan dari algoritma K-Nearest Neighbors dalam mengklasifikasikan sentimen dari sebuah tweet yang berkaitan dengan vaksin booster sebagai syarat mudik. Pada penelitian ini menggunakan metodologi Cross-Industry Standard Process For Data Mining (CRISP-DM). Dengan melewati beberapa tahapan seperti tahapan preprocessing dan pemodelan. Jumlah data yang didapat 2384 dengan sentimen positif 1970 tweet dan sentimen negatif 414 tweet. Berdasarkan hasil yang didapat dari tahapan pemodelan dengan menggunakan algoritma K-Nearest Neighbors pemodelan split data perbandingan 70:30 menggunakan k=9, maka accuracy yang dihasilkan sebesar 85.17%, precision 85.51% dan recall 98.82%

2.1.3 Cholid Fadilah Hasri, Debby Alita(2022)⁶

Di penelitian ini dengan judul “Penerapan Metode Naïve Bayes Classifier dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter” dilakukan pada 2 juni 2022. Pemerintahan Indonesia telah memberlakukan PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) di sejumlah daerah Indonesia sepanjang tahun 2021. Larangan untuk berkumpul dan

⁵ Andreansyah and Grace Gata, “Analisis Sentimen Untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN,” *KRESNA: Jurnal Riset Dan Pengabdian Masyarakat* 2, no. 2 (2022): 193–201, <https://doi.org/10.36080/jk.v2i2.48>.

⁶ Cholid Fadilah Hasri and Debby Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)* 3, no. 2 (2022): 145–60, <http://jim.teknokrat.ac.id/index.php/informatika>.

berkerumun selama PPKM telah diberlakukan untuk mengurangi mobilitas masyarakat agar mencegah terjadinya penyebaran COVID-19. Berita ini pun cepat menyebar di berbagai platform media sosial salah satunya adalah media sosial twitter. Setidaknya sejumlah masyarakat berbondong-bondong mengungkapkan pendapat mereka mengenai kebijakan-kebijakan yang telah dibuat oleh pemerintah pusat dan daerah tentang PPKM di tahun 2021. Dengan memanfaatkan teknik Text Mining metode klasifikasi, akan diketahui suatu sentimen bernilai positif, netral atau negatif. Salah satu algoritma yang banyak digunakan dalam analisis sentimen adalah metode klasifikasi Naïve Bayes. Penelitian ini menggunakan metode Naïve Bayes Classifier (NBC). Hasil penelitian menunjukkan bahwa metode Naïve Bayes tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 81.07% Untuk memastikan hasil penelitian ini, dilakukan juga pengujian dengan Support Vector Machine yang menghasilkan nilai akurasi sebesar 79.96%.

2.1.4 Debby Alita, RB Ali Shodiqin(2023)⁷

Penelitian ini dilakukan pada tahun 2023 dengan judul “Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine” Pemberian vaksin di Indonesia saat ini sudah mencapai tahapan vaksin booster, berbagai jenis vaksin telah diberikan kepada masyarakat Indonesia dari vaksin Sinovac, astrazeneca, Sinopharm, Moderna, Pfizer, dsb. Tidak sedikit masyarakat Indonesia yang menggunakan beberapa jenis vaksin yang ditawarkan hingga vaksin booster akan tetapi ada beberapa masyarakat yang berpendapat masih terinfeksi virus covid ini dengan gejala berat. Pendapat lain juga terdapat kemudian terdapat vaksin Pada tahun 2019 masyarakat dihebohkan dengan adanya virus baru dari Wuhan, China yaitu virus corona atau disebut COVID-19 (Coronavirus Disease 2019). Pemerintah.

Pembahasan yang telah dilakukan Pada Penelitian analisis sentimen terhadap vaksin covid-19 dapat ditarik kesimpulan dari penelitian ini. Dimana kesimpulan dari penelitian yang telah dilakukan antara lain analisis sentimen tweet menunjukkan jumlah sentimen positif sebanyak 790 tweet dan jumlah

⁷ Debby Alita and R B Ali Shodiqin, “Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine,” *Journal of Artificial Intelligence and Technology Information (JAITI)* 1, no. 1 (2023): 1–12.

sentimen negatif sebanyak 361 tweet. Maka dikatakan bahwa lebih banyak pengguna Twitter yang memberikan respon positif terhadap kegiatan Vaksin Covid-19. Penerapan metode *Naive Bayes Classifier* menunjukkan hasil performance vector dengan nilai accuracy 72.88%, precision 43.49% dan recall 54.95%. Penerapan metode *Support Vector Machine* menunjukkan hasil pengujian dengan nilai Accuracy 77.88%, Precision 75.00%, dan Recall 77.00%. Metode *Naive Bayes Classifier* memiliki kemampuan yang lebih baik dalam menganalisis sentimen Twitter dengan nilai rata-rata performance 57.10% dibandingkan dengan *Support Vector Machine* hanya mencapai rata-rata performance 40,77%.

2.1.5 Dadang Iskandar Mulyana, Nesti Lutfianti(2023)⁸

Saat ini banyak negara yang sedang memperbincangkan kemungkinan terjadinya resesi di negaranya, Indonesia termasuk diantaranya. Sinyal terjadinya kemungkinan resesi dapat terlihat dari mulai melemahnya aktivitas ekonomi sektor riil pada negara tersebut. Dengan adanya isu berita tersebut, menuai banyak pendapat dan opini dari netizen melalui platform media sosial twitter, mereka saling berkomentar terhadap adanya isu berita resesi 2023. penelitian ini dengan judul “Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023” Dilakukan pada tahun 2023 Tujuan dari penelitian analisis sentimen ini adalah untuk memahami dan mengetahui opini masyarakat atas isu berita resesi pada 2023 pada media sosial twitter apakah opini tersebut masuk dalam kelas sentimen positif atau kelas sentimen negatif serta mengetahui hasil akurasi persentase dari setiap kelas.

Pengujian dilakukan menggunakan model klasifikasi *support vector machine (SVM)* dan beberapa metode seperti *data crawling*, *pre-processing*, dan pembobotan kata. Berdasarkan hasil pengujian model didapatkan hasil akurasi 98,67% nilai *recall* positif sebesar 99.33%, nilai *recall* negatif sebesar 98.00%, nilai *precision* positif sebesar 98.03% dan nilai *precision* negatif sebesar 99.32%. Data latih sebanyak 300 data sebanyak 149 data terprediksi positif dan 1 data

⁸ Jurnal Sisfotenika, “Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023 Analysis Sentiment Using the SVM Algorithm in Netizen Responses to News of the 2023 Recession” 13, no. 1 (2023): 53–64.

terprediksi negatif. Untuk hasil prediksi sentimen negatif terdapat 147 data dan 3 data terprediksi positif. Dari proses sentimen analisis dengan 667 data uji, dihasilkan prediksi sebanyak 373 Sentimen Positif dan 294 Sentimen Negatif. Dalam penelitian ini sumber data yang digunakan hanya berasal dari satu sumber media social, yakni media social twitter. Hal ini dapat menjadi rujukan bagi penelitian selanjutnya dengan menggunakan sumber data media sosial lainnya yang berbeda.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Tahun	Judul	Kata Kunci	Metode Analisis
1	Ezza Ramadhanta Machmud Razaq, Deden Witarsyah Jacob, Faqih Hamami	2021	<i>Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid 19 Pada Media Sosial Twitter menggunakan Perbandingan Algoritma Klasifikasi</i>	<i>“Pembelajaran Online Saat Pandemi” 1548 data relevan yang digunakan</i>	Ketiga algoritma tersebut didapatkan akurasi tertinggi yaitu 60.80% menggunakan algoritma klasifikasi K-Nearest Neighbor (K-NN). Hasil akurasi yang didapat menggunakan ekstraksi fitur TF-IDF sebesar 60.80%, presisi (specificity) 40.73%, recall (sensitivity) 72%, dan F1-Measure 38.43%. Analisis sentimen yang didapat pada penelitian ini menunjukkan bahwa pengguna twitter di Indonesia lebih banyak memberikan komentar netral.
2.	Andreansyah, Grace Gata	2022	<i>Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik</i>	<i>“Vaksin Booster Sebagai Syarat Mudik” Data Yang digunakan 2384</i>	Jumlah dataset yang digunakan 2384 data dengan sentimen positif 1970 data dan sentimen negatif 414 data. Maka

			<i>2022 Menggunakan Algoritma KNN</i>		dapat disimpulkan bahwa metode algoritma K-Nearest Neighbors mendapatkan nilai akurasi sebesar 85.17% saat menggunakan model pengujian Split Data dengan perbandingan 70:30 dan saat model pengujian Cross Validation dengan k=9 mendapatkan nilai akurasi sebesar 84.4%.
3.	Cholid Fadilah Hasri, Debby Alita	2022	<i>Penerapan Metode Naïve Bayes Classifier dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter</i>	<i>“Dampak Virus Corona” 1104 tweet di ambil dari Twitter</i>	Hasil penelitian menunjukkan bahwa metode Naïve Bayes tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 81.07% Untuk memastikan hasil penelitian ini, dilakukan juga pengujian dengan Support Vector Machine yang menghasilkan nilai akurasi sebesar 79.96%.

4	Debby Alita, RB Ali Shodiqin	2023	<i>Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine</i>	<i>“Vaksin Covid-19” 1151 data di ambil dari Twitter</i>	Penerapan metode Naive Bayes Classifier menunjukkan hasil performance vector dengan nilai accuracy 72.88%, precision 43.49% dan recall 54.95%. Penerapan metode Support Vector Machine menunjukkan hasil pengujian dengan nilai Accuracy 77.88%, Precision 75.00%, dan Recall 7.70%. Metode Naive Bayes Classifier memiliki kemampuan yang lebih baik dalam menganalisis sentimen Twitter dengan nilai rata-rata performance 57.10% dibandingkan dengan Support Vector Machine hanya mencapai rata-rata performance 40,77%.
---	------------------------------	------	---	--	---

5.	Dadang Iskandar Mulyana, Nesti Lutfianti	2023	<i>Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023</i>	<i>“Resesi 2023” Data diambil dari twitter Menggunakan 2000 data</i>	Pengujian dilakukan menggunakan model klasifikasi support vector machine (SVM) dan beberapa metode seperti data crawling, pre-processing, dan pembobotan kata. Berdasarkan hasil pengujian model didapatkan hasil akurasi 98,67% nilai Recall Positif sebesar 99.33%, nilai Recall Negatif sebesar 98.00%, nilai Precision Positif sebesar 98.03% dan nilai Precision Negatif sebesar 99.32%.
----	--	------	---	--	---

6	Faiz Aunullah	2023	Perbandingan Metode <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> Pada Analisis Sentimen Terhadap Isu Resesi 2023 Pada Media Sosial Twiter Dan Komentar Youtube	<i>“Resesi 2023 dan Resesi”</i> Menggunakan 5571 data Yang diambil dari tanggal 10 Oktober 2022 sampai 3 Januari 2023.	Membandingkan dengan metode <i>Naïve Bayes</i> dengan metode <i>Support Vector Machine</i> dengan hasil <i>accuracy</i> pada <i>Naïve Bayes</i> sebesar 0.729 dan <i>Support Vector Machine</i> sebesar 0.800. Pada penelitian ini metode <i>Support Vector Machine</i> mempunyai tingkat <i>accuracy</i> yang lebih baik dibandingkan metode <i>Naïve Bayes</i>
---	---------------	------	--	---	--

2.2 Kajian Teori

2.2.1 Resesi

Resesi adalah sebuah istilah yang digunakan untuk menggambarkan keadaan dimana perputaran ekonomi suatu negara berubah menjadi lambat atau buruk. Perputaran ekonomi yang melambat ini bisa berlangsung cukup lama bahkan tahunan akibat dari pertumbuhan produk domestik bruto (PDB) suatu negara menurun selama dua kuartal dan berlangsung secara terus menerus.⁹

2.2.2 Media Sosial

Media Sosial adalah sebuah media online, dan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Jejaring sosial merupakan situs dimana setiap orang bisa membuat web pribadi, kemudian terhubung dengan teman teman untuk berbagi informasi dan berkomunikasi. Jejaring sosial terbesar antara lain Facebook, Myspace, dan Twitter.¹⁰

2.2.3 Analisis Sentimen

Analisis sentimen atau opinion data mining adalah deteksi perilaku perilaku terhadap objek atau orang. Analisis sentimen dapat digunakan untuk menerima persentase positif, negatif serta netral terhadap seseorang, perusahaan, atau produk. Kelas pada analisis sentimen dibagi menjadi tiga kelas yaitu kelas positif, negatif, dan netral. Atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif.¹¹

Analisis sentimen bertujuan untuk melakukan evaluasi terhadap emosi, sikap, pendapat, dan penilaian yang disampaikan oleh seseorang terhadap suatu permasalahan. Ulasan tersebut menyebabkan beberapa penelitian terutama pada

⁹ “Resesi Adalah: Pengertian, Penyebab, Dampak, Dan Cara Menghadapinya,” n.d., <https://www.bfi.co.id/id/blog/resesi-adalah-pengertian-penyebab-dampak-dan-cara-menghadapinya>.

¹⁰ Anang Sugeng Cahyono, “Anang Sugeng Cahyono, Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat Di Indonesia,” n.d., 140–57.

¹¹ Fransiska Vina Sari, “ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD . ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI” 10, no. 2 (2019): 681–86.

review produk didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining

2.2.4 Netlytic

Netlytic merupakan web penganalisis teks dan jaringan sosial berbasis cloud yang dapat secara otomatis merangkum data tekstual dan menemukan jaringan komunikasi dari postingan media sosial yang dapat diakses publik dan juga sebagai salah satu aplikasi yang membantu penelitian jaringan komunikasi.¹²

2.2.5 Naïve Bayes

Naïve Bayes Classifier (NBC) adalah metode klasifikasi yang berdasarkan teorema bayes. Metode ini menggunakan probabilitas dan statistika yang pertama kali dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes, yaitu suatu metode untuk memprediksi masa depan berdasarkan pengalaman di masa sebelumnya sehingga metode ini sering dikenal dengan teorema bayes.¹³

Naive bayes banyak digunakan dalam Teknik klasifikasi terutama pada twitter menggunakan beberapa metode seperti *Unigram Naïve bayes*, *Multinomial Naïve Bayes*, dan *Maximum Entropy Classification*. Fitur utama klasifikasi *Naïve Bayes* untuk mendapatkan hipotesis yang kuat dari suatu kondisi atau peristiwa. Pada *Teorema Bayes*, bila terdapat dua kejadian yang terpisah (misalkan X dan H), maka *Teorema Bayes* dirumuskan sebagai berikut.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad 2.1$$

Dimana X merupakan tupel data hasil pengujian dari suatu set data yang telah ditentukan masuk ke dalam kelas tertentu. H merupakan suatu hipotesis yang menentukan X masuk ke dalam kelas C. $P(H|X)$ merupakan peluang atau probabilitas X yang merupakan tuple atau bukti yang diperoleh pada saat observasi masuk ke dalam C, dengan kata lain mencari probabilitas X dimiliki

¹² Gema Maulida Pangestu, Merry Fridha Tripalupi, and Judhi Hari Wibowo, "Analisis Jaringan Komunikasi # KesetaraanGender Sebagai Aktivitas Digital Di Twitter Dalam Meningkatkan Kesadaran Tentang Kesetaraan Gender," n.d.

¹³ Magister Teknik Informatika and Universitas Amikom, "ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING" 3, no. 2 (2019): 179–84.

oleh C. $P(H|X)$ merupakan probabilitas posterior, H dikondisikan pada X. Sebaliknya $P(H)$ merupakan probabilitas prior, dan probabilitas sebelumnya. Kemudian $P(X|H)$ merupakan probabilitas posterior dimana X di kondisikan pada H. Sedangkan $P(X)$ merupakan probabilitas sebelumnya dari X.

2.2.6 Support Vector Machine

Support Vector Machine (SVM) dikembangkan pada tahun 1992 oleh Vladimir Vapnik bersama dengan kedua rekannya, Bernhard Boser dan Isabelle Guyon. SVM dikembangkan dari teori struktural risk minimization. Sebagai suatu Teknik klasifikasi efisien untuk masalah non linear, *Support Vector Machine (SVM)* juga dikenal sebagai teknik pembelajaran machine learning setelah pembelajaran mesin sebelumnya yang disebut *Neural Network (NN)*. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperline terbaik yang berfungsi sebagai pemisah kelas pada input space. SVM dapat menjamin kemampuan generalisasi tinggi untuk data-data yang akan datang.¹⁴

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti *Support Vector Classification*) dan regresi (*Support Vector Regression*). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear.

¹⁴ Zidna Alhaq et al., "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *Journal of Information System Management (JOISM)* 3, no. 2 (2021): 44–49, <https://doi.org/10.24076/joism.2021v3i2.558>.

2.2.7 Confusion Matrix

Confusion Matrix merupakan tabel yang menyatakan klasifikasi jumlah data uji benar dan jumlah data uji yang salah.¹⁵ Contoh *confusion matrix* ditunjukkan pada Tabel 2.2.

Tabel 2.2 Confusion Matrix

		Tabel Prediksi		
		Positif	Negatif	Netral
Kelas Sebenarnya	Positif	TP	FN	FNT
	Negatif	FP	TN	FNT
	Netral	FNT	FNT	TNT

Keterangan :

TP = True Positif

TN = True Negatif

TNR = True Netral

FP = Positif (True Positif, False Netral, False Negatif)

FN = Negatif (True Negatif, False Positif, False Netral)

FNR = Netral (True Netral, False Positif, False Negatif)

Rumus confusion matrix untuk menghitung *accuracy*, *precision* dan *recall* adalah sebagai berikut.

$$Akurasi = \frac{TP+TN+TNR}{P+N+NR} \times 100\% \quad 2.2$$

$$Precision = \frac{TP}{TP+FN} \quad 2.3$$

$$Recall = \frac{TP}{TP+FP} \quad 2.4$$

¹⁵ Dwi Normawati and Surya Allit Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)* 5, no. 2 (2021): 697–711, <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>.

BAB 3 METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

3.1.1 Waktu

Penelitian ini dilakukan dalam 6 bulan, terhitung dari bulan Oktober 2022 sampai Maret 2023. Perancangan Kegiatan penelitian disajikan pada Tabel 3.1.

Tabel 3.1 Timeline Penelitian

No	Kegiatan	2022		2023		
		November	Desember	Januari	Februari	Maret
1	Analisis Kebutuhan					
2	Pengumpulan Data					
3	Pelabelan Data					
4	Validasi Data					
5	Desain Sistem					
6	Training Data					
7	Implementasi Sistem					
8	Pengujian sistem dan evaluasi					
9	Laporan					

3.1.2 Tempat

Penelitian ini akan dilakukan di Universitas Darussalam Gontor Ponorogo

3.2 Analisis Kebutuhan

Alat dan kebutuhan sistem yang digunakan dalam pembuatan penelitian ini adalah :

3.2.1 Perangkat Keras :

1. Perangkat Laptop :
 - AMD A9-9425 RADEON R5, 5 COMPUTE 2c + 3G (2 cpus),~3.1ghz
 - Layar 14”.
 - Ram 8GB DDR4
 - HDD 1000GB
 - SSD 256 GB

3.2.2 Perangkat Lunak :

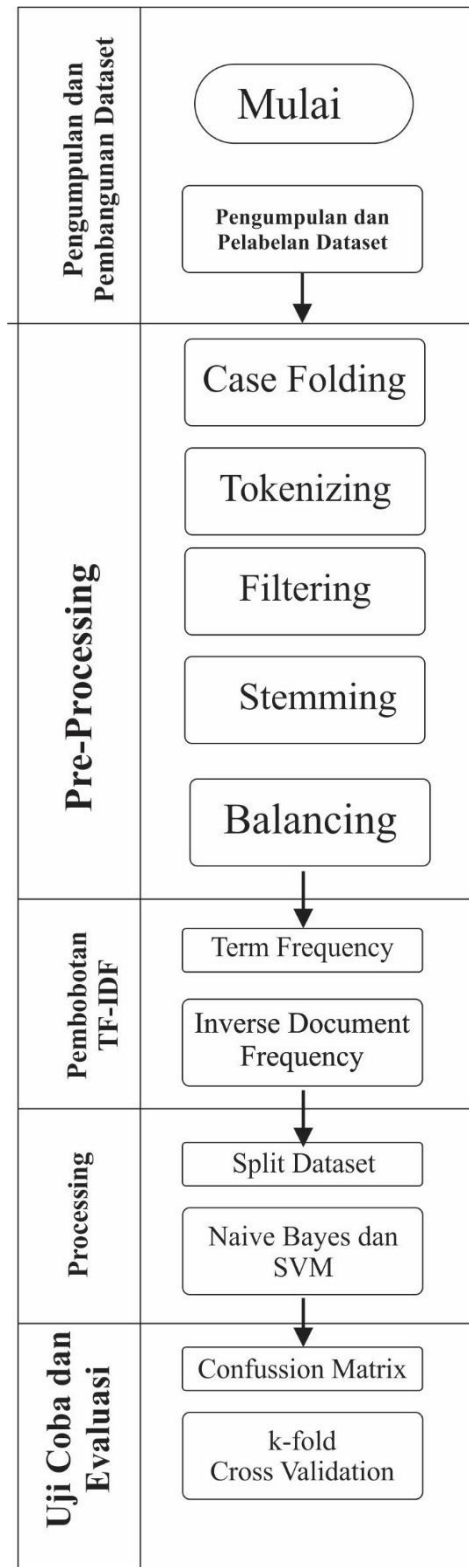
1. Windows 11 pro
2. Microsoft office professional plus 2019
3. Google Collabs
4. Python 3.7
5. Mendeley

3.3 Perancangan Sistem

Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan sentiment atau pendapat masyarakat terhadap isu resesi 2023 pada kumpulan tweets dan komentar pada youtube dalam opini Positif, Negatif, dan Netral. Proses klasifikasi dilakukan menggunakan dua algoritma yaitu algoritma *Naïve Bayes* dan *Support Vector Machine*. Model diimplementasikan untuk menganalisa sentiment pengguna twitter dan youtube dengan satu kata kunci yaitu resesi yang diambil pada tanggal 10 Oktober 2022 sampai 3 Januari 2023.

Pada perancangan sistem klasifikasi ini memiliki beberapa tahapan tertera pada Gambar 3.1. Sistem ini dirancang menggunakan 5 tahapan yaitu :

1. Pengumpulan dan Pelabelan Dataset
2. *Pre-Processing*
3. Pembobotan TF-IDF
4. *Processing*
5. Uji Coba dan Evaluasi



Gambar 3.1 Tahapan Perancangan Sistem

3.3.1 Pengumpulan dan Pembangunan Dataset

Teknik yang digunakan untuk pengumpulan data dengan menggunakan crawling data yang diaplikasikan pada sebuah website yaitu dengan alamat dengan menggunakan website dengan alamat <https://netlytic.org/index.php> memudahkan peneliti untuk mengambil data pada media sosial. Untuk pengambilan data dari twitter dengan jangka waktu 7 hari terakhir dan komentar pada youtube. Jumlah maksimal pengambilan data menggunakan website ini adalah 2500 data. Pengumpulan data dimulai dari tanggal 10 Oktober hingga 3 Januari dengan menggunakan kata kunci “Resesi” dan “Resesi 2023” sehingga menghasilkan data sebanyak 5502 data dengan pembagian 2002 data dari twitter dan 3500 data dari video youtube dengan judul “Gelapnya Perekonomian Indonesia di Tahun 2023”

Pelabelan dilakukan secara manual dengan dibagi menjadi tiga kelas yaitu kelas positif, negatif, dan netral. Data positif merupakan komentar yang setuju akan tidak terjadinya resesi. Sedangkan negatif merupakan komentar yang setuju terjadinya resesi dan berkomentar tidak baik. Netral untuk komentar yang tidak relevan dengan pembahasan. Data yang telah diberi label lalu di validasi oleh Almas Rizqullah seorang guru Bahasa Indonesia di SMP Muhammadiyah 7 Surabaya.

3.3.2 Pre-Processing

Karena data yang dikumpulkan merupakan data instruktur maka diperlukan tahapan *Pre-Processing* untuk menjadikan data tersebut menjadi terstruktur dengan melakukan *Pre-Processing*. Tahapan *Pre-Processing* ditunjukan pada gambar 3.1, dalam proses ini kumpulan data dilakukan *Pre-Processing* untuk melakukan transformasi data. Tahapan ini dilakukan penyiapan data dalam teks karena masih memiliki beberapa permasalahan seperti data duplikat, noise dll. Setelah tahapan *Pre-Processing* kita dapat melanjutkan ke tahap *Processing* yaitu menerapkan metode yang akan kita gunakan dalam penelitian ini Dalam *Pre-Processing* terdapat beberapa tahapan yaitu :

1. *Case Folding*

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.¹⁶

2. *Tokenizing*

Tokenizing atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

3. *Filtering*

Filtering adalah tahap mengambil kata-kata penting dari hasil term. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-word. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan seterusnya.

4. *Stemming*

Stemming merupakan proses mengubah kata menjadi kata dasarnya. *Stemming* dilakukan menggunakan library Sastrawi.¹⁷

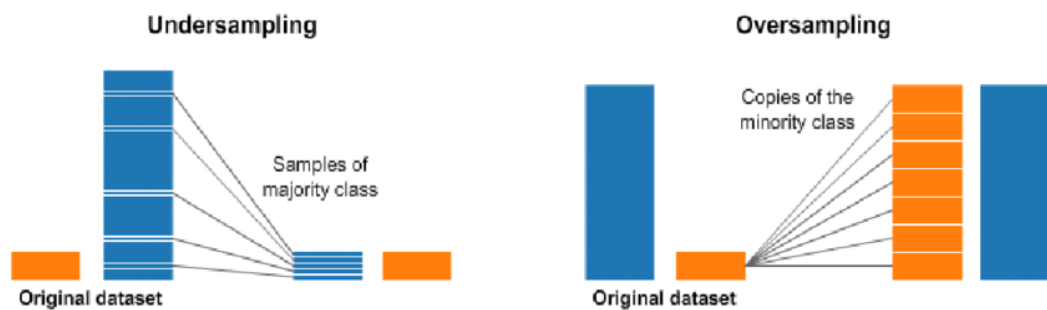
5. *Balancing*

Kumpulan data yang tidak seimbang adalah kumpulan data yang memiliki kemiringan parah dalam pendistribusian kelas seperti contoh 1:100 atau 1:1000 di kelas minoritas ke kelas mayoritas. Biasanya dalam dataset dapat mempengaruhi banyak mesin pembelajaran algoritma. Maka dari itu perlu di lakukan penyeimbangan data. Untuk membuat data menjadi seimbang terdapat dua metode yaitu *Random Undersampling* dan *Random Oversampling*.

¹⁶ Ade Hendriani and Susy Katarina Sianturi, “Sentiment Masyarakat Terhadap Virus COVID-19 Pada Instagram Menggunakan Algoritma Naïve Bayes Classifier,” *J-SAKTI (Jurnal Sains Komputer Dan Informatika)* 5, no. 1 (2021): 443–52, <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/336/315>.

¹⁷ Fakhrana Kurnia Sutrisno and Kemas Muslim Lhaksmana, “Analisis Sentimen Destinasi Wisata Kuliner Di Twitter Menggunakan TF-IDF Dan Complement Naïve Bayes Pada Dataset Tidak Seimbang,” *Agustus* 8, no. 4 (2021): 4115.

Random Oversampling bertujuan untuk meningkatkan sampel kelas minoritas sampai sama dengan mayoritas lain dengan menduplikasi secara acak sampel dari kelas minoritas. Sedangkan *Random Undersampling* yaitu menghasilkan sub sampel acak dari instance mayoritas. Berikut contoh dari *Random Oversampling* dan *Random Undersampling*



Gambar 3.2 Contoh Undersampling dan Oversampling

Pada penelitian ini menggunakan *Teknik Random Oversampling* untuk membuat data pada kelas menjadi seimbang. *Random Oversampling* merupakan metode pembangkitan data minoritas sebanyak data mayoritas.¹⁸

3.3.3 Pembobotan TF-IDF

Ekstraksi fitur dilakukan untuk memperoleh informasi penting dari teks. Pada penelitian ini, ekstraksi fitur yang dilakukan menggunakan metode *TF-IDF*.

TF-IDF atau *Term Frequency-Inverse Document Frequency* merupakan sebuah perhitungan yang digunakan untuk menghitung bobot kata dalam suatu dokumen. Lebih besar frekuensi kemunculan sebuah kata pada suatu dokumen, kata tersebut lebih memungkinkan untuk menjadi kata yang mewakili dokumen tersebut. Kata pada suatu atau beberapa dokumen memiliki nilai *TF-IDF* yang besar daripada kata umum seperti *Pre-Processing*. Karena kata kata tersebut termasuk kedalam *Stopwords* yang kemudian dihapus untuk meningkatkan kinerja program.¹⁹

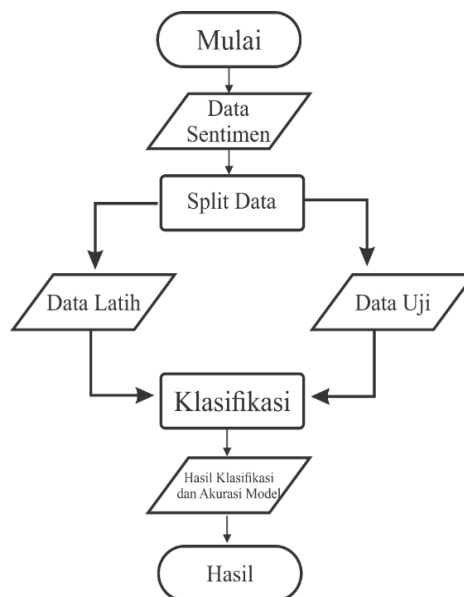
¹⁸ Ade Yuni Triyanto and Retno Kusumaningrum, "Implementasi Teknik Sampling Untuk Mengatasi Imbalanced Data Pada Penentuan Status Gizi Balita Dengan Menggunakan Learning Vector Quantization," *JURNAL IPTEKKOM : Jurnal Ilmu Pengetahuan & Teknologi Informasi* 19, no. 1 (2017): 39, <https://doi.org/10.33164/iptekkom.19.1.2017.39-50>.

¹⁹ Ni Komang Widayanti, I Ketut Gede Darma Putra, and Ni Kadek Dwi Rusjyanthi, "Seleksi Fitur Bobot Kata Dengan Metode TFIDF Untuk Ringkasan Bahasa Indonesia," *Jurnal Ilmiah*

Metode *TF-IDF* dapat digunakan untuk berbagai tujuan seperti klasifikasi dokumen, penentuan topik, dan pengambilan informasi. Sebagai contoh, dalam klasifikasi dokumen kata-kata yang memiliki bobot *TF-IDF* yang tinggi dalam kategori tertentu dapat digunakan sebagai fitur untuk membedakan dokumen dari kategori tersebut dengan dokumen dari kategori lain.

3.3.4 Processing

Pada tahap ini data yang telah melalui proses labelling data *TF-IDF* akan dibagi menjadi dua yaitu data training (latih) dan data testing (uji). Metode yang digunakan untuk mengklasifikasi data tersebut adalah metode *Naïve Bayes*, dan *Support Vector Machine* sesuai dengan Gambar 3.3.



Gambar 3.3 Flowchart Klasifikasi²⁰

Gambar 3.4 menjelaskan bahwa tahap pertama membaca seluruh data yang ada setelah dilakukan *Text Pre-Processing*. Data akan dibagi menjadi dua yaitu data training (latih) dan data testing (uji). Data training akan menjadi referensi bagi ketiga metode untuk mengklasifikasikan data, data testing untuk menguji hasil dari pembelajaran metode tersebut. Hasil yang

Merpati (*Menara Penelitian Akademika Teknologi Informasi*) 6, no. 2 (2018): 119, <https://doi.org/10.24843/jim.2018.v06.i02.p06>.

²⁰ Amalia Indranandita, Budi Susanto, and Antonius Rahmat, "Sistem Klasifikasi Dan Pencarian Jurnal Dengan Menggunakan Metode Naive Bayes Dan Vector Space Model," *Jurnal Informatika* 4, no. 2 (2011), <https://doi.org/10.21460/inf.2008.42.48>.

didapatkan berupa model yang dapat digunakan untuk mengklasifikasi dan nilai akurasinya.²¹

3.3.5 Uji Coba dan Evaluasi

Evaluasi pada penelitian ini menggunakan metode *Confusion Matrix*. Metode ini berguna untuk menilai kinerja model klasifikasi yang dibuat. Dalam menilai kinerja suatu model yang dibuat *Confusion Matrix* membandingkan hasil klasifikasi yang diperoleh dengan hasil klasifikasi yang seharusnya pada dasarnya metode ini membandingkan hasil label testing dengan label training.

Actual	Negatif	True Negatif	False Negatif	False Negatif
	Netral	False Netral	True Netral	False Netral
	Positif	False Postif	False Postif	True Positif
		Negatif	Netral	Positif
		Prediction		

Gambar 3.4 3x3 Confusion Matrix²²

Confusion matrix yang digunakan pada penelitian ini adalah dimensi 3x3. Dimensi tersebut sesuai dengan tiga kelas pada penelitian ini yaitu positif, negatif, dan netral Gambar 3.5 menunjukkan persamaan matrix 3x3 yang digunakan.²³

Perhitungan dimensi 3x3 sesuai gambar 3.5 adalah sebagai berikut :

$$Akurasi = \frac{TP+TN+TNR}{P+N+NR} \times 100\% \quad 3.1$$

²¹ Aziz Mustafa et al., “Analisis Sentimen Opini Masyarakat Terhadap Virus Omicron Di Indonesia Menggunakan Metode Naïve Bayes Abstrak” 7, no. 2 (2023).

²² Srikanth Nagineni and Rajesh M. Hegde, “On line client-wise cohort set selection for speaker verification using iterative normalization of confusion matrices”, *European Signal Processing Conference*, no. 1 (2010), pp. 576–80.

²³ Ali Imron, “KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” n.d.

Keterangan :

TP = True Positif

TN = True Negatif

TNR = True Netral

P = Positif (*True Positif, False Netral, False Negatif*)

N = Negatif (*True Negatif, False Positif, False Netral*)

NR = Netral (*True Netral, False Positif, False Negatif*)

Tabel 3.2 K-fold Validation²⁴

No	Pembagian Dataset									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10
Keterangan										
	Nomor Pengujian									
	Data Test									
	Data Training									

²⁴ Dihin Muriyatmoko, Triana Harmini, and Maulana Kemal Ardiansyah, "Sentiment Analysis Covid-19 Vaccination on Twitter Social Media Using Naïve Bayes Method," *Procedia of Engineering and Life Science* 2, no. 2 (2021), <https://doi.org/10.21070/pels.v2i0.1144>.

Setelah data uji diklasifikasikan di uji dan validasi menggunakan *K-fold Validation*. Yang berfungsi untuk mengacak data uji untuk menemukan rata rata akurasi dari model yang dibuat. Pada penelitian ini akan dilakukan sebanyak sepuluh kali dengan urutan bertingkat seperti Tabel 3.2 dimana setiap kelas terdiri dari 10% dari keseluruhan data.²⁵

²⁵ Huzain Azis, Farniwati Fattah, and Pratiwi Putri, “Performa Klasifikasi K-NN Dan Cross-Validation Pada Data Pasien Pengidap Penyakit Jantung” 12, no. 2 (2020): 81–86.

BAB 4 HASIL DAN PEMBAHASAN

Pada tahapan ini akan dibahas hasil dan pembahasan tentang implementasi yang telah dilakukan. Pada penelitian ini peneliti melakukan penelitian tentang klasifikasi sentimen masyarakat terhadap isu resesi 2023 pada kumpulan tweet dan komentar pada youtube dalam opini positif, negatif, dan netral. Parameter yang digunakan untuk mengukur tingkat keberhasilan pada model adalah nilai akurasi. Nilai akurasi model dapat ditentukan dengan melakukan pengujian menggunakan data testing pada proses *Confusion Matrix*.

4.1 Hasil

Pada tahapan ini berisikan tentang hasil dari proses pengumpulan data dan pembangunan dataset, *Pre-Processing* dan Pembobotan TF-IDF

4.1.1. Pengumpulan Data dan Pembangunan Dataset

Data yang digunakan dalam penelitian penerapan metode *naïve bayes*, *support vector machine*, dan *random forest* terhadap isu resesi 2023. Penelitian ini menggunakan data dari twitter dan youtube yang diambil secara acak dimulai pada tanggal 3 Oktober 2022 sampai 3 Januari 2023 berjumlah 5502 data.

Pelabelan terhadap 5502 data yang diteliti dilakukan secara manual dalam tiga kelas sentiment positif, negatif, dan netral dengan rincian pada Tabel 4.1.

Tabel 4.1 Jumlah Data Per Kategori Kelas

No	Kelas	Jumlah
1	Positif	2741
2	Negatif	1752
3	Netral	1009
Total		5502

Pada Tabel 4.1 menunjukkan bahwa data pada kelas positif berjumlah 2741 data, pada kelas negatif berjumlah 1752 data, dan pada kelas netral berjumlah 1009 data.



4.1.2. Pre-Processing

Tahapan tahapan pada *Text Pre-processing* yang pertama terdapat *Case Folding, Tokenizing, Filtering, Stemming, dan Balancing*.

a) *Case Folding*,

Tahapan pada *Text Pre-processing* yang bertujuan untuk membuat text menjadi konsisten dan teratur. Yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil. Karakter selain huruf dihilangkan dan dianggap delimiter. Contoh perubahan dapat dilihat pada Tabel 4.2


Tabel 4.2 Contoh Tahapan *Case Folding*

Sebelum Case Folding	Setelah Case Folding	Sentimen
@Askrlfess @catuaries Aku jadi kamu bakalan minta alat masak lebih lengkap dan belanjain segala macam bahan masak, kalo perlu sampe dibayarin kursus masak, trus open PO jualan makanan. Manfaatin selama belum putus. Apalagi mau resesi, skill masak tetep pasti laku karena semua butuh makan.	@askrlfess @catuaries aku jadi kamu bakalan minta alat masak lebih lengkap dan belanjain segala macam bahan masak, kalo perlu sampe dibayarin kursus masak, trus open po jualan makanan. manfaatin selama belum putus. apalagi mau resesi, skill masak tetep pasti laku karena semua butuh makan.	Positif
ANJAI 	anjai 	Netral
Alah bcrit lah,bilang saja utang sudah banyak kluar	alah bcrit lah,bilang saja utang sudah banyak kluar	Negatif

b) *Tokenizing*

Tokenizing, adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya agar bisa masuk ke tahapan *filtering* untuk dianalisa. Fungsi split pada python digunakan untuk memisahkan kata. Contoh perubahan dapat dilihat pada Tabel 4.3

Tabel 4.3 Contoh Hasil *Tokenizing*


Sebelum Tokenizing	Setelah Tokenizing	Sentimen
@askrlfess @catuaries aku jadi kamu bakalan minta alat masak lebih lengkap dan belanjain segala macam bahan masak, kalo perlu sampe dibayarin kursus masak, trus open po jualan makanan. manfaatin selama belum putus. apalagi mau resesi, skill masak tetep pasti laku karena semua butuh makan.	['alat', 'masak', 'lengkap', 'belanjain', 'macem', 'bahan', 'masak', 'kalo', 'sampe', 'dibayarin', 'kursus', 'masak', 'trus', 'open', 'po', 'jual', 'makan', 'manfaatin', 'putus', 'resesi', 'skill', 'masak', 'tetep', 'laku', 'butuh', 'makan']	Positif
anjai 	['anja', '"', '"', '"', '"']	Netral
alah bcrit lah,bilang saja utang sudah banyak keluar	['alah', 'bcrit', 'lahbilang', 'utang', 'keluar']	Negatif

c) *Filtering*

Filtering, adalah tahap mengambil kata-kata penting dari hasil term. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang

penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-word*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan seterusnya. Contoh perubahan dapat dilihat dari Tabel 4.4

Tabel 4.4 Contoh Hasil *Filtering*

Sebelum Filtering	Setelah Filtering	Sentimen
@askrlfess @catuaries aku jadi kamu bakalan minta alat masak lebih lengkap dan belanjain segala macam bahan masak, kalo perlu sampe dibayarin kursus masak, trus open po jualan makanan. manfaatin selama belum putus. apalagi mau reresi, skill masak tetep pasti laku karena semua butuh makan.	alat masak lengkap belanjain macam bahan masak kalo sampe dibayarin kursus masak trus open po jual makan manfaatin putus reresi skill masak tetep laku butuh makan	Positif
anjai 	anjai	Netral
alah bcrit lah,bilang saja utang sudah banyak keluar	alah bcrit lahbilang utang keluar	Negatif

d) *Stemming*

Stemming adalah proses menghilangkan infleksi kata. *Stemming* digunakan untuk mengkonversi kata yang telah melewati tahap *filtering* ke kata dasar yang tidak memiliki imbuhan. Misalnya kata “memiliki”, ”dimiliki” menjadi kata “milik” atau proses mengubah kata menjadi kata dasarnya. *Stemming* dilakukan menggunakan library python yang

telah disediakan yaitu Sastrawi.²⁶ Contoh perubahan dapat dilihat pada Tabel 4.5.

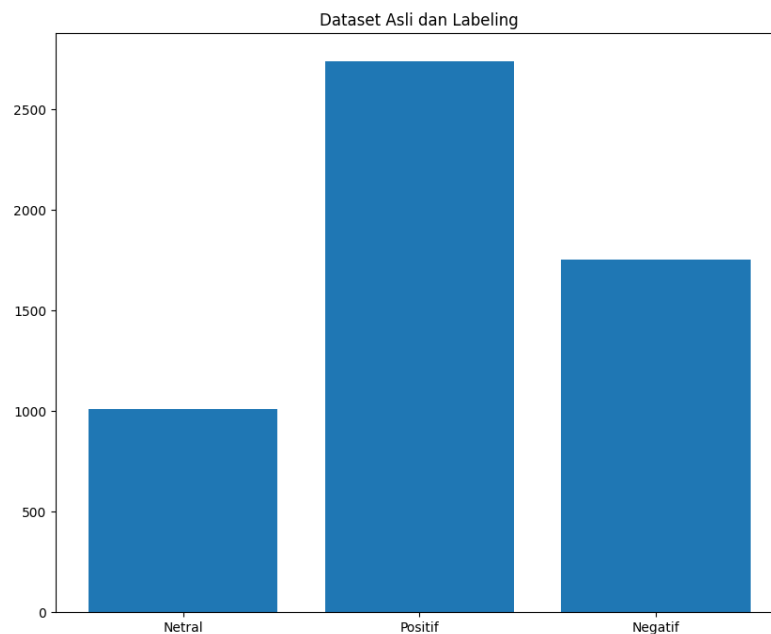
Tabel 4.5 Hasil Stemming

Sebelum Stemming	Setelah Stemming	Sentimen
@askrlfess @catuaries aku jadi kamu bakalan minta alat masak lebih lengkap dan belanjain segala macam bahan masak, kalo perlu sampe dibayarin kursus masak, trus open po jualan makanan. manfaatin selama belum putus. apalagi mau resesi, skill masak tetep pasti laku karena semua butuh makan.	alat masak lengkap belanjain macam bahan masak kalo sampe dibayarin kursus masak trus open po jual makan manfaatin putus resesi skill masak tetep laku butuh makan	Positif
anjai	anja	Netral
alah bcrit lah,bilang saja utang sudah banyak keluar	alah bcrit lahbilang utang keluar	Negatif

²⁶ Kurnia Sutrisno and Muslim Lhaksana, "Analisis Sentimen Destinasi Wisata Kuliner di Twitter Menggunakan TF-IDF dan Complement Naïve Bayes pada Dataset Tidak Seimbang".

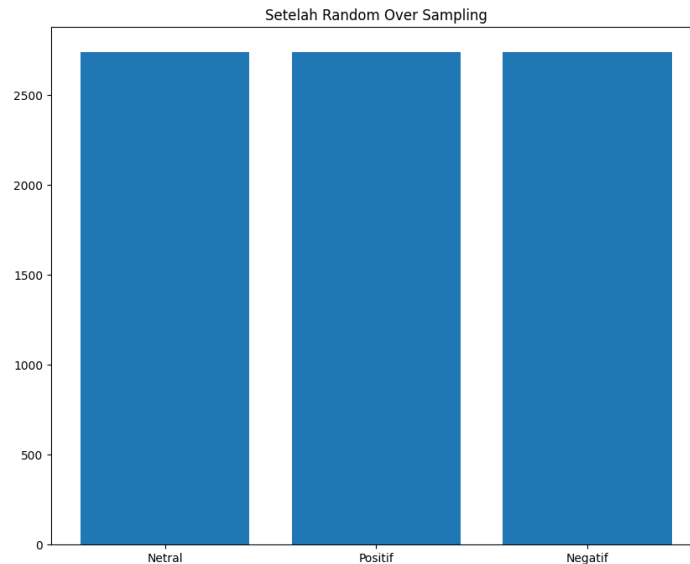
e) *Balancing*

Klasifikasi merupakan pemodelan yang memasukkan data kedalam suatu kelas. Jumlah data yang masuk kedalam tiap kelas merupakan distribusi .dalam pendistribusian data terkadang terdapat perbedaan jumlah data pada kelas tersebut atau sering disebut *imbalance* data. Pada penelitian ini menggunakan Teknik *Random Oversampling* untuk menjadikan data tersebut menjadi *balance*. *Random Oversampling* merupakan metode pembangkitan data minoritas sebanyak data mayoritas. Berikut hasil dari sebelum dilakukan *Random oversampling* dan sesudah melakukan *Random Oversampling* data pada gambar 4.1.²⁷



Gambar 4.1 Sebelum *Random Oversampling*

²⁷ Akhmad Miftahusalam et al., “Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer”, *Seminar Nasional Official Statistics*, vol. 2022, no. 1 (2022), pp. 563–72.



Gambar 4.2 Setelah *Random Oversampling*

Pada Gambar 4.1 menjelaskan data keseluruhan pada kelas tidak seimbang sehingga perlu dilakukanya teknik balancing data untuk membuat data tersebut menjadi seimbang dengan hasil pada Gambar 4.2. data menjadi 2741 data yang mana sesuai dengan data yang paling banyak pada penelitian ini yaitu data positif.

4.1.3. Pembobotan TF-IDF

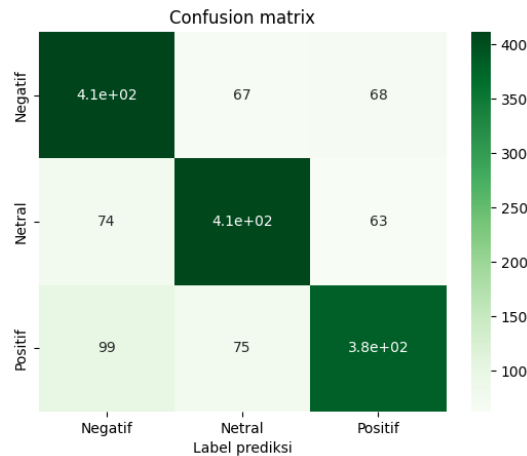
Setelah melakukan tahapan *pre-processing* tahapan selanjutnya adalah pembobotan kata atau ekstraksi fitur untuk mengeluarkan informasi yang digunakan. Pada penelitian ini menggunakan model dari metode *TF-IDF* dengan pembagian perkata. *TF-IDF* pada python menggunakan library *tfidftransformer*.²⁸

Fungsi dari tahap ini adalah mengubah data teks menjadi data numerik supaya bisa melakukan perhitungan dan untuk menghitung bobot dari setiap kata. Semakin besar bobot maka data tersebut semakin penting. pembobotan *TF-IDF* juga berguna untuk penyaringan data karena kata yang memiliki bobot akan diproses pada tahap selanjutnya. Sedangkan yang bernilai 0 maka tidak akan proses. Adapun hasil pembobotan *TF-IDF* terdapat pada Gambar 4.2.

²⁸ Kurnia Sutrisno and Muslim Lhaksmana, "Analisis Sentimen Destinasi Wisata Kuliner di Twitter Menggunakan TF-IDF dan Complement Naïve Bayes pada Dataset Tidak Seimbang".

4.2 Hasil Klasifikasi Naïve Bayes

Tahap ini menggunakan model klasifikasi dengan metode *Naïve Bayes* menggunakan data yang sudah di siapkan. Pemodelan ini menggunakan data training yang sudah berlabel. Pada penelitian ini menggunakan 20% data testing dan 80% untuk data training. Pengklasifikasian ini menghasilkan *Confusion Matrix* ditampilkan pada Gambar 4.3



Gambar 4.3 Naïve Bayes Confussion Matrix

Pada Gambar 4.3 menampilkan pemodelan *Naïve Bayes* terdapat kelas positif, negatif, dan netral yang masih tersebar kedalam kelas kelas lain yang tidak sesuai pada kelasnya hingga pada pemodelan menggunakan metode *Naïve Bayes* Menghasilkan nilai akurasi sebesar 0.729 dengan rincian pada tabel 4.6

Tabel 4.6 Precision, Recall Dan F1-Score Naïve Bayes

Kelas	Precision	Recall	F1-Score
Negatif	0.70	0.75	0.73
Netral	0.74	0.75	0.74
Positif	0.74	0.69	0.71
AVG	0.73	0.73	0.73

Tabel 4.6 menunjukkan nilai *precision*, *recall* dan *f1-score* pada model *Naïve Bayes* dengan rata-rata *precision* 0.73, *recall* 0.73, dan *F1-Score* 0.73. Setelah tahapan evaluasi model akan di validasi menggunakan *K-Fold Cross Validation* dengan membagi data menjadi beberapa bagian dengan rasio yang sama namun dengan sampel data yang berbeda pada setiap bagian.

Dengan melakukan pelatihan menggunakan data training dan pengujian terhadap data testing. Pada setiap percobaan mengambil satu segmen yang berbeda. Pengujian model klasifikasi dilakukan menggunakan *10-Fold Cross Validation* yang hasilnya dapat dilihat pada Tabel 4.7.

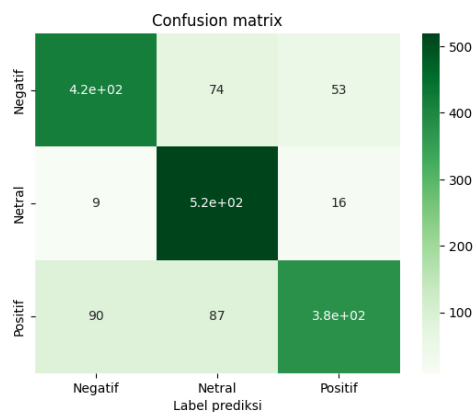
Tabel 4.7 Rincian *K- Fold Naïve Bayes*

K-Fold	<i>Naive Bayes</i>
1	0.69148936
2	0.75075988
3	0.72796353
4	0.69908815
5	0.72036474
6	0.72796353
7	0.72492401
8	0.72948328
9	0.70928463
10	0.71689498
AVG	0.719821607842593

Pada tabel 4.7 menunjukan bahwa hasil rata rata menggunakan metode *Naïve Bayes* sebesar **0.719821607842593** nilai tertinggi adalah 0.72492401, dan nilai terendahnya adalah 0.69148936.

4.3 Hasil Klasifikasi Support Vector Machine

Tahap selanjutnya menggunakan model klasifikasi dengan metode *Support Vector Machine (SVM)* menggunakan data yang sudah disiapkan. Pemodelan ini menggunakan data training yang sudah berlabel. Pada penelitian ini menggunakan 20% data testing dan 80% untuk data training. Pengklasifikasian ini menghasilkan *Confusion Matrix* ditampilkan pada Gambar 4.4.



Gambar 4.4 SVM Confusion Matrix

Gambar 4.4 menampilkan pemodelan *Support Vector Machine* (SVM) terdapat kelas positif, negatif, dan netral yang masih tersebar kedalam kelas kelas lain yang tidak sesuai pada kelasnya hingga pada pemodelan menggunakan metode *Support Vector Machine* (SVM) menghasilkan nilai akurasi sebesar 0.800 dengan rincian pada tabel 4.8.

Tabel 4.8 Precision, Recall Dan F1-Score SVM

Kelas	Precision	Recall	F1-Score
Negatif	0.81	0.77	0.79
Netral	0.76	0.95	0.85
Positif	0.85	0.68	0.75
AVG	0.81	0.80	0.80

Tabel 4.8 menunjukkan nilai *Precision*, *Recall* dan *f1-Score* pada model *Support Vector Machine* (SVM) dengan rata-rata *precision* 0.81, *recall* 0.80, dan *F1-Score* 0.80. Setelah tahapan evaluasi model akan di validasi menggunakan *K-Fold Cross Validation* dengan membagi data menjadi beberapa bagian dengan rasio yang sama namun dengan sampel data yang berbeda pada setiap bagian.

Dengan melakukan pelatihan menggunakan data training dan pengujian terhadap data testing. Pada setiap percobaan mengambil satu segmen yang berbeda. Pengujian model klasifikasi dilakukan menggunakan *10-Fold Cross Validation* yang hasilnya dapat dilihat pada Tabel 4.9.

Tabel 4.9 Rincian K- Fold Support Vector Machine

K-Fold	SVM
1	0.75835866
2	0.79787234
3	0.7781155
4	0.78571429
5	0.78723404
6	0.79635258
7	0.7887538
8	0.80699088
9	0.78234399
10	0.77321157
AVG	0.7854947652819992

Pada tabel 4.9 menunjukan bahwa hasil rata rata menggunakan metode *Support Vector Machine (SVM)* sebesar **0.7854947652819992**. Nilai tertinggi adalah 0.80699088, dan nilai terendahnya adalah 0.78234399.

4.4 Hasil Perbandingan

Berikut Hasil Perbandingan metode pada Tabel 4.12.

Tabel 4.10 Hasil Perbandingan

No	Metode	Accuracy
1	<i>Naïve Bayes</i>	0.729
2	<i>Support Vector Machine</i>	0.800

Pada Tabel 4.12 membuktikan bahwasanya pada penelitian ini metode *Support Vector Machine (SVM)* Memiliki tingkat akurasi yang lebih baik dibandingkan metode *Naïve Bayes* dengan akurasi sebesar 0.800.

4.5 Implementasi Sistem

Setelah memperoleh hasil dari perbandingan ketiga metode tersebut, langkah selanjutnya adalah analisa sentimen data tweet pada tanggal 10 Oktober 2022 sampai tanggal 3 januari 2023 menggunakan 2 kata kunci yang berjumlah 2544 data riil dengan rincian pada tabel 4.11

Tabel 4.11 Jumlah Data Per Kata Kunci

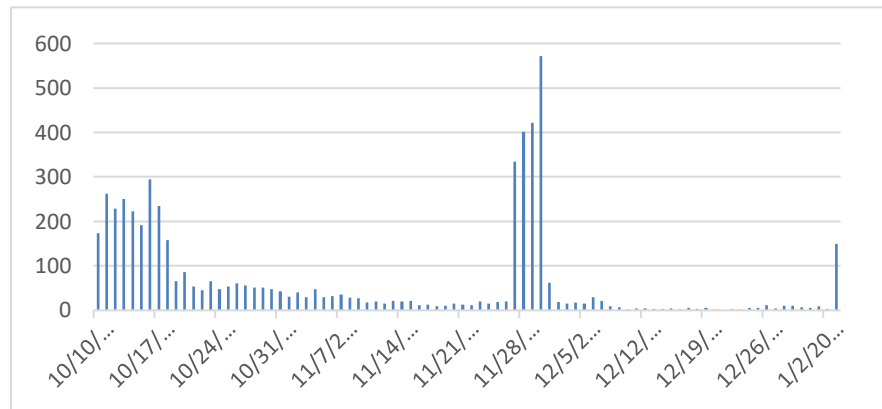
No	Kata Kunci	Jumlah
1	Resesi 2023	325
2	Resesi	2192

Model implementasi pada data tersebut dengan hasil berupa tweet yang telah terklasifikasi kedalam 3 kelas yaitu positif, negatif, dan netral.

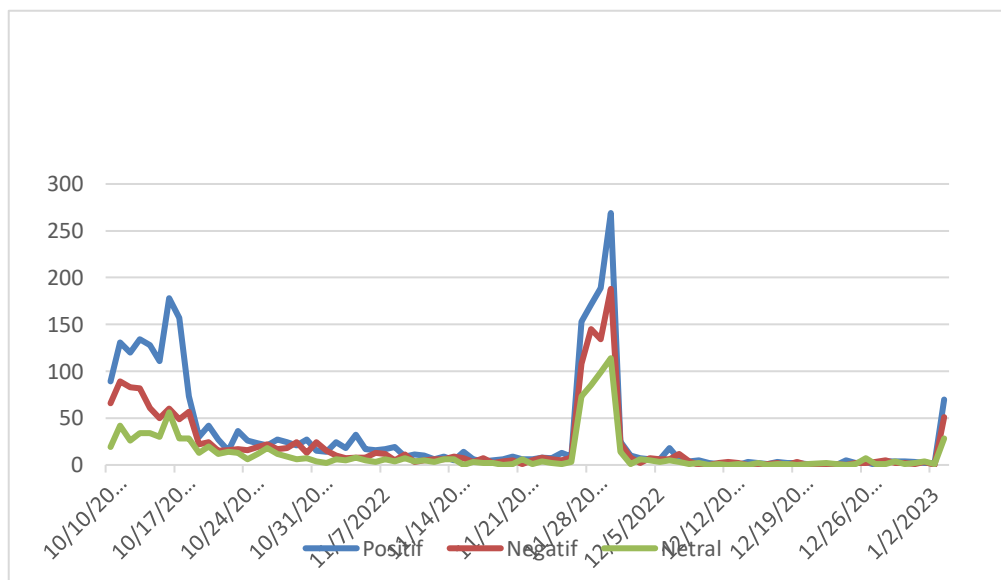
4.5.1 Hasil Analisis Penyebaran Tren Data

Penyebaran tren data tweet per hari pada ketiga kata kunci pada tanggal 10 Oktober 2022 sampai dengan 3 Januari 2023 dapat dilihat pada gambar 4.5

Gambar 4.5 Tren Sentimen Data Per Tanggal



Data yang berhasil dikumpulkan dengan rentan tanggal tersebut sebanyak 5505 data. Hasil klasifikasi sentimen positif, negatif, dan netral per hari pada tanggal tersebut terinci pada Gambar 4.6.



Gambar 4.6 Tren Sentimen Data Per Hari

Pada Gambar 4.6 dapat dilihat bahwa tren jumlah data bersentimen positif paling tinggi terjadi pada tanggal 30 November 2022. Di barengi dengan tingginya tren data pada tanggal tersebut.

Tabel 4.12 Tren Sentimen Data Per Kata Kunci

Kata Kunci	Negatif	Netral	Positif
Resesi 2023	18%	23%	60%
Resesi	33%	20%	47%
AVG	25%	21%	53%

Untuk persentase sentimen pengguna *twitter* terhadap tiga kata kunci yaitu Resesi 2023, Ekonomi Dunia dan Resesi Dunia pada tanggal 10 Oktober 2022 sampai dengan 3 Januari 2023 didominasi oleh sentimen positif dengan persentase 53% seperti pada Tabel 4.12. Tren Jumlah data paling tinggi terjadi pada tanggal 30 November 2022 yang di barengi dengan tingginya tren sentimen positif.

BAB 5 PENUTUP

5.1 Kesimpulan

Kesimpulan dari penelitian perbandingan metode *Naïve Bayes* dan *Support Vector Machine* pada analisis sentimen masyarakat terhadap isu resesi 2023 pada media sosial twitter dan komentar youtube sebagai berikut :

1. Model yang dibuat dapat mengklasifikasikan sentimen masyarakat terhadap isu resesi 2023 pada media sosial twitter dan komentar pada youtube dalam opini positif, negatif, dan netral menggunakan metode *Naïve Bayes* dan *Support Vector Machine*.
2. Model yang dibuat menggunakan metode *Naïve Bayes* menghasilkan nilai akurasi sebesar 0.729, dan menggunakan metode *Support Vector Machine* menghasilkan nilai akurasi sebesar 0.800.
3. Pada penelitian ini metode *Support Vector Machine (SVM)* Memiliki tingkat akurasi yang lebih baik dibandingkan metode *Naïve Bayes* dengan akurasi sebesar 0.800.
4. Hasil pelabelan data menampilkan bahwa sentimen positif lebih banyak daripada sentiment netral dan negatif. Menunjukkan banyaknya masyarakat yang percaya jika resesi 2023 tidak terjadi.

5.2 Saran

Dari hasil penelitian ini, peneliti menyadari bahwa penelitian ini masih banyak kekurangan, sehingga perlu dikembangkan lebih lanjut agar mendapatkan hasil yang lebih maksimal. Beberapa saran yang dapat diambil adalah sebagai berikut :

1. Menambahkan lebih banyak jumlah data, karena dataset akan sangat mempengaruhi hasil dari model dan akurasi pada klasifikasi.
2. Memodifikasi pada tahapan pre-processing sehingga dapat meningkatkan hasil akurasi.
3. Mencoba menggunakan metode lain untuk mendapatkan hasil akurasi yang berbeda pada klasifikasi.

DAFTAR PUSTAKA

- Alhaq, Zidna, Ali Mustopa, Sri Mulyatun, And Joko Dwi Santoso. “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter.” *Journal Of Information System Management (JOISM)* 3, No. 2 (2021): 44–49. <https://doi.org/10.24076/Joism.2021v3i2.558>.
- Alita, Debby, And R B Ali Shodiqin. “Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine.” *Journal Of Artificial Intelligence And Technology Information (JAITI)* 1, No. 1 (2023): 1–12.
- Andreansyah, And Grace Gata. “Analisis Sentimen Untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN.” *KRESNA: Jurnal Riset Dan Pengabdian Masyarakat* 2, No. 2 (2022): 193–201. <https://doi.org/10.36080/Jk.V2i2.48>.
- Azis, Huzain, Farniwati Fattah, And Pratiwi Putri. “Performa Klasifikasi K-NN Dan Cross-Validation Pada Data Pasien Pengidap Penyakit Jantung” 12, No. 2 (2020): 81–86.
- Cahyono, Anang Sugeng. “Anang Sugeng Cahyono, Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat Di Indonesia,” N.D., 140–57.
- Hasri, Cholid Fadilah, And Debby Alita. “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter.” *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)* 3, No. 2 (2022): 145–60. <http://jim.teknokrat.ac.id/index.php/informatika>.
- Hendriani, Ade, And Susy Katarina Sianturi. “Sentiment Masyarakat Terhadap Virus COVID-19 Pada Instagram Menggunakan Algoritma Naïve Bayes Classifier.” *J-SAKTI (Jurnal Sains Komputer Dan Informatika)* 5, No. 1 (2021): 443–52. <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/336/315>.
- Imron, Ali. “KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” N.D.
- Indranandita, Amalia, Budi Susanto, And Antonius Rahmat. “Sistem Klasifikasi

- Dan Pencarian Jurnal Dengan Menggunakan Metode Naive Bayes Dan Vector Space Model.” *Jurnal Informatika* 4, No. 2 (2011). <https://doi.org/10.21460/Inf.2008.42.48>.
- Informatika, Magister Teknik, And Universitas Amikom. “ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING” 3, No. 2 (2019): 179–84.
- Kurnia Sutrisno, Fakhrana, And Kemas Muslim Lhaksana. “Analisis Sentimen Destinasi Wisata Kuliner Di Twitter Menggunakan TF-IDF Dan Complement Naïve Bayes Pada Dataset Tidak Seimbang.” *Agustus* 8, No. 4 (2021): 4115.
- Miraza, Bachtiar Hassan. “Seputar Resesi Dan Depresi.” *Jurnal Ekonomi Kiat* 30, No. 2 (2019): 11–13. <https://journal.uir.ac.id/index.php/kiat>.
- Muriyatmoko, Dihin, Triana Harmini, And Maulana Kemal Ardiansyah. “Sentiment Analysis Covid-19 Vaccination On Twitter Social Media Using Naïve Bayes Method.” *Procedia Of Engineering And Life Science* 2, No. 2 (2021). <https://doi.org/10.21070/Pels.V2i0.1144>.
- Mustafa, Aziz, Triana Harmini, Angga Fahri Setiawan, Nur Aini, And Shofiya Asy. “Analisis Sentimen Opini Masyarakat Terhadap Virus Omicron Di Indonesia Menggunakan Metode Naïve Bayes Abstrak” 7, No. 2 (2023).
- Normawati, Dwi, And Surya Allit Prayogi. “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter.” *Jurnal Sains Komputer & Informatika (J-SAKTI)* 5, No. 2 (2021): 697–711. <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>.
- Pangestu, Gema Maulida, Merry Fridha Tripalupi, And Judhi Hari Wibowo. “Analisis Jaringan Komunikasi # Kesetaraangender Sebagai Aktivitas Digital Di Twitter Dalam Meningkatkan Kesadaran Tentang Kesetaraan Gender,” N.D.
- PRIJAMBODO, BAMBANG. “Resesi Dunia 2023, Seberapa Parah?” www.kompas.id, 2023. <https://www.kompas.id/baca/opini/2023/02/19/resesi-dunia-2023-seberapa-parah>.
- Ramadhanta Machmud Razaq, Ezza, Jacob Deden Witarsyah, And Faqih Hamami. “Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online

- Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi.” *E-Proceeding Of Engineering* 8, No. 5 (2021): 9000–9006.
- “Resesi Adalah: Pengertian, Penyebab, Dampak, Dan Cara Menghadapinya,” N.D. <https://www.bfi.co.id/blog/resesi-adalah-pengertian-penyebab-dampak-dan-cara-menghadapinya>.
- Sari, Fransiska Vina. “ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD . ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI” 10, No. 2 (2019): 681–86.
- Sisfotenika, Jurnal. “Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023 Analysis Sentiment Using The SVM Algorithm In Netizen Responses To News Of The 2023 Recession” 13, No. 1 (2023): 53–64.
- “Sri Mulyani Bilang AS Lolos Dari Resesi 2023, RI Gimana?” Accessed April 18, 2023. <https://www.cnbcindonesia.com/news/20230303100106-4-418558/sri-mulyani-bilang-as-lolos-dari-resesi-2023-ri-gimana>.
- Triyanto, Ade Yuni, And Retno Kusumaningrum. “Implementasi Teknik Sampling Untuk Mengatasi Imbalanced Data Pada Penentuan Status Gizi Balita Dengan Menggunakan Learning Vector Quantization.” *JURNAL IPTEKKOM : Jurnal Ilmu Pengetahuan & Teknologi Informasi* 19, No. 1 (2017): 39. <https://doi.org/10.33164/iptekkom.19.1.2017.39-50>.
- Widyasanti, Ni Komang, I Ketut Gede Darma Putra, And Ni Kadek Dwi Rusjayanthi. “Seleksi Fitur Bobot Kata Dengan Metode TFIDF Untuk Ringkasan Bahasa Indonesia.” *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)* 6, No. 2 (2018): 119. <https://doi.org/10.24843/Jim.2018.V06.I02.P06>.