

**ANALISIS SENTIMEN ULASAN PRODUK TOKOPEDIA
MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER :
STUDI KASUS TOKO PENGRAJIN.COM**

LAPORAN PROYEK PROFESSIONAL



MUHAMMAD AULIA

5200411482

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS TEKNOLOGI YOGYAKARTA
YOGYAKARTA**

2023

LAPORAN PROYEK PROFESSIONAL

**ANALISIS SENTIMEN ULASAN PRODUK TOKOPEDIA
MENGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER
: STUDI KASUS TOKO PENGRAJIN.COM**

disusun oleh

MUHAMMAD AULIA

5200411482



Dosen Pembimbing

Dr. Ir. Arief Hermawan, ST., MT., IPU
NIK. 111094017

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini, saya

Nama : Muhammad Aulia

NPM : 5200411482

Program Studi : Informatika

Program : Sarjana

Fakultas : Sains & Teknologi

menyatakan bahwa tugas akhir dengan judul Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier : Studi Kasus Toko Pengrajin.Com ini adalah karya ilmiah asli saya dan belum pernah dipublikasikan oleh orang lain, kecuali yang tertulis sebagai acuan dalam naskah ini dan disebutkan dalam daftar pustaka. Apabila di kemudian hari, karya saya disinyalir bukan merupakan karya asli saya, maka saya bersedia menerima konsekuensi apa yang diberikan Program Studi Informatika Fakultas Sains dan Teknologi Universitas Teknologi Yogyakarta.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta

Pada tanggal : 10 Juni 2023

Yang menyatakan

Muhammad Aulia

ABSTRAK

Penelitian ini menggunakan metode Naïve Bayes Classifier dengan algoritma SMOTE & Imblearn. Naïve Bayes Classifier adalah algoritma klasifikasi berdasarkan teorema Bayes dengan asumsi fitur independen. Metode SMOTE & Imblearn digunakan untuk mengatasi ketidakseimbangan kelas pada dataset ulasan produk. SMOTE membuat sampel sintetis ulasan negatif untuk menjaga keseimbangan kelas, sementara Imblearn digunakan untuk meningkatkan jumlah sampel pada kelas minoritas dan meningkatkan akurasi analisis sentimen.. Tujuan dari penelitian ini adalah untuk menganalisis sentimen ulasan produk di Tokopedia dengan menggunakan algoritma Naïve Bayes Classifier dan melihat pengaruh penerapan metode SMOTE & Imblearn dalam meningkatkan akurasi analisis sentimen. Melalui penelitian ini, diharapkan dapat memperoleh pemahaman yang lebih baik tentang pendapat dan tanggapan masyarakat terhadap produk yang ditawarkan oleh Toko Pengrajin.com. Selain itu, penelitian ini juga bertujuan untuk memberikan masukan berharga kepada perusahaan atau instansi terkait untuk meningkatkan kualitas produk dan layanan mereka berdasarkan respon masyarakat yang terungkap melalui analisis sentimen.

Hasil penelitian ini menunjukkan bahwa penggunaan algoritma Naïve Bayes Classifier dengan penerapan metode SMOTE & Imblearn mampu memberikan akurasi analisis sentimen yang lebih tinggi. Dengan menggunakan metode tersebut, penelitian mencapai tingkat akurasi sebesar 87% dalam mengklasifikasikan ulasan produk di Tokopedia. Sementara itu, tanpa menggunakan metode SMOTE & Imblearn, akurasi yang diperoleh sebesar 87%. Temuan ini menunjukkan bahwa penggunaan algoritma Naïve Bayes Classifier dalam analisis sentimen terhadap ulasan produk di Tokopedia, khususnya pada Toko Pengrajin.com, sangat efektif. Selain itu, penerapan metode SMOTE & Imblearn juga memberikan kontribusi signifikan dalam meningkatkan akurasi analisis sentimen dengan memperbaiki ketidakseimbangan kelas dalam dataset. Hasil penelitian ini memberikan informasi berharga bagi perusahaan atau instansi terkait untuk meningkatkan kualitas produk dan layanan mereka berdasarkan respon masyarakat yang terungkap melalui analisis sentimen.

Kata Kunci : *Naïve Bayes Classifier*, SMOTE&Imblearn, Analisis Sentimen, Pengrajin.com, Tokopedia

KATA PENGANTAR

Puji syukur dipanjatkan atas kehadiran Allah SWT, karena dengan limpahan karunia-Nya penulis dapat menyelesaikan Proyek Informatika dengan judul Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier : Studi Kasus Toko Pengrajin.Com

Penyusunan Laporan Proyek Informatika ini dapat diselesaikan tidak lepas dari segala bantuan, bimbingan, dorongan dan doa dari berbagai pihak, yang pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada :

- a. Bapak Dr. Bambang Moertono Setiawan, M.M., Akt., CA. selaku Rektor Universitas Teknologi Yogyakarta.
- a. Bapak Sutarman, M.Kom., Ph.D. selaku Dekan Fakultas Sains & Teknologi Universitas Teknologi Yogyakarta.
- b. Ibu Dr. Enny Itje Sela S.Si, M.Kom., selaku ketua Program Studi Informatika Fakultas Sains & Teknologi Universitas Teknologi Yogyakarta.
- c. Bapak Arief Hermawan, Dr., S.T., M.T. selaku Dosen Pembimbing yang telah berkenan membimbing dan mengarahkan penulis dalam penyelesaian Proyek Profesional.
- d. Kedua orang tua yang selalu support dan memberikan dukungan serta doa demi kelancaran Proyek Profesional.
- e. Cahyo, Ilham, dan Rousydi yang telah memberikan dukungan, membantu perbaikan error, serta memberikan semangat, motivasi dan juga membantu dalam penulisan serta penyusunan laporan Proyek Profesional.
- f. Seluruh pihak yang tidak dapat disebutkan satu persatu, yang telah mendukung dan memberikan masukan kepada penulis selama pengerjaan laporan Proyek Profesional.

Akhir kata, penulis menyadari bahwa sepenuhnya akan terbatasnya pengetahuan penyusunan sehingga tidak menutup kemungkinan, jika ada kesalahan serta kekurangan dalam penyusunan laporan Proyek Informatika, untuk itu sumbang

saran dari pembaca sangat diharapkan sebagai bahan pelajaran berharga di masa yang akan datang.

Yogyakarta, 10 Juni 2023

Penulis

DAFTAR ISI

ABSTRAK	iii
KATA PENGANTAR.....	iv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penelitian.....	3
BAB II KAJIAN HASIL PENELITIAN DAN LANDASAN TEORI.....	6
2.1 Kajian Hasil Penelitian	6
2.2 Landasan Teori	12
2.2.1 E-Commerce	12
2.2.2 Tokopedia.....	13
2.2.3 Pengrajin.Com.....	14
2.2.4 Scraping	14
2.2.5 Analisis Sentimen	14
2.2.6 Naïve Bayes Classifier	15
2.2.7 Text Preprocessing	16
2.2.8 TF-IDF	16
2.2.9 Natural Language Processing (NLP)	17
2.2.10 Confusion Matrix	17
2.2.11 Akurasi	18
2.2.12 Python	19
2.2.13 Visualisasi	23
BAB III METODE PENELITIAN	25
3.1 Bahan/Data	25
3.1.1 Data yang Diperoleh	25
3.1.2 Prosedur Pengumpulan Data	25
3.2 Aturan Bisnis (Bussiness Rule).....	27
3.2.1 Analisis Sistem Saat ini.....	27
3.2.2 Kelemahan/Kekurangan Sistem Saat ini	27
3.3 Tahapan Penelitian	28
BAB IV ANALISIS DAN PERANCANGAN SISTEM	29
4.1 Analisis Sistem.....	29
4.1.1 Analisis Sistem yang Berjalan	29
4.1.2 Analisis Sistem yang Diusulkan	29
4.2 Desain Sistem.....	32
4.2.1 Perancangan Logik.....	32
BAB V IMPLEMENTASI DAN HASIL SERTA PEMBAHASAN	37
5.1 Implementasi	37
5.1.1 Implementasi Proses Scraping	37
5.1.2 Implementasi Proses Preprocessing	38

5.1.3 Implementasi Proses Machine Learning	40
5.1.4 Implementasi Proses Prediction	42
5.2 Hasil	43
5.2.1 Implementasi Halaman Scraping	44
5.2.2 Implementasi Halaman Preprocessing	44
5.2.3 Implementasi TF-IDF	47
5.2.4 Training dan Testing	48
5.2.5 Implementasi Halaman Machine Learning	48
5.2.6 Implementasi Halaman Prediction	51
BAB VI PENUTUP	54
6.1 Simpulan	54
6.2 Saran.....	54
DAFTAR PUSTAKA	56

DAFTAR GAMBAR

Gambar 3. 1	Data Ulasan Tokopedia.....	25
Gambar 3. 2	Ulasan Toko Pengrajin.Com.....	26
Gambar 3. 3	Tahapan Penelitian	28
Gambar 4. 1	Diagram Konteks	33
Gambar 4. 2	Diagram Jenjang	34
Gambar 4. 3	Diagram Aliran Data Level 1	34
Gambar 4. 4	Diagram Aliran Data Level 2	35
Gambar 4. 5	Flowchart Sistem	36
Gambar 5. 1	Script Scraping	38
Gambar 5. 2	Script Preprocessing	40
Gambar 5. 3	Script Machine Learning	42
Gambar 5. 4	Script Prediction	43
Gambar 5. 5	Tampilan Scraping.....	44
Gambar 5. 6	Tampilan Preprocessing	46
Gambar 5. 7	Tampilan Machine Learning tanpa SMOTE	49
Gambar 5. 8	Tampilan Machine Learning dengan SMOTE	51
Gambar 5. 9	Tampilan Prediction Sentimen Negatif	52
Gambar 5. 10	Tampilan Prediction Sentimen Positif	53

DAFTAR TABEL

Tabel 2. 1. Sumber Pustaka Primer	10
Tabel 2. 2. Confusion Matrix	18
Tabel 5. 1 Hasil TF	47
Tabel 5. 2 Hasil TF-IDF	48

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini masyarakat di Indonesia bahkan dunia lebih senang berbelanja barang-barang yang mereka inginkan menggunakan online shop. Salah satu online shop yang sering digunakan oleh masyarakat Indonesia adalah Tokopedia. Hal tersebut terjadi dikarenakan masyarakat menilai menggunakan online shop akan menjadi lebih mudah dan praktis serta tidak membuang-buang waktu untuk pergi ke toko secara langsung. Selain itu, ada juga faktor yang mempengaruhi hal tersebut adalah barang yang dibeli pelanggan lebih murah karena banyaknya promo yang diberikan oleh toko online. Sehingga akan lebih banyak menarik pelanggan online shop untuk mengunjungi toko online tersebut.

Potensi pasar yang terdapat dalam karakteristik perilaku penggunaan internet di Indonesia telah membuat banyak online shop bermunculan dengan berbagai macam produk dan jasa yang ditawarkan kepada pembeli, sehingga persaingan antar online shop untuk dapat menarik perhatian konsumen membeli semakin ketat. Tingginya tingkat persaingan antar online shop menyebabkan konsumen bertindak selektif dalam menentukan keputusan pembelian karena berbelanja melalui internet memiliki banyak risiko. Kondisi tersebut menuntut para pelaku online shop untuk dapat memahami kebutuhan dan keinginan yang semakin kompleks.

Analisis sentimen atau opinion mining merupakan analisis terhadap suatu pendapat seseorang pada sesuatu hal tertentu yang dituangkan ke dalam sebuah pendapat, perkataan, atau komentar untuk memberikan masukan tersendiri bagi pihak yang sedang diberikan opini. Pada proses penelitian yang akan dilakukan ini, peneliti akan membagi pelabelan data ke dalam tanggapan positif dan negatif untuk ditentukan sebagai bahan penelitian. Pada penelitian yang akan dijalankan terkait dengan referensi penelitian sebelumnya, yaitu "Implementasi Algoritma Naïve Bayes Terhadap Analisis

Sentimen Opini Film pada Twitter", berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90% dengan rincian nilai precision 92%, recall 90%, dan f-measure 90% (Ratnawati, F., 2018).

Dalam studi kasus ini, akan dilakukan analisis sentimen terhadap ulasan produk di Tokopedia dengan menggunakan algoritma Naïve Bayes Classifier. Fokus penelitian ini adalah toko Pengrajin.com, yang merupakan salah satu toko online yang terdaftar di Tokopedia. Melalui analisis sentimen, diharapkan dapat memperoleh wawasan tentang pendapat pengguna terhadap produk-produk yang ditawarkan oleh Pengrajin.com di platform Tokopedia.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

- a. Bagaimana melakukan analisis sentimen terhadap ulasan produk di Tokopedia menggunakan algoritma Naïve Bayes Classifier?
- b. Bagaimana mengidentifikasi dan membedakan sentimen positif dan negatif dalam ulasan pelanggan terhadap produk-produk di toko Pengrajin.com di Tokopedia?
- c. Bagaimana menentukan kualitas produk di toko Pengrajin.com berdasarkan hasil analisis sentimen positif dan negatif dari ulasan pelanggan menggunakan metode algoritma Naïve Bayes Classifier?
- d. Apa saja wawasan yang dapat diperoleh dari analisis sentimen terhadap ulasan pelanggan terkait produk-produk Pengrajin.com di Tokopedia?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

- a. Penelitian akan menggunakan data ulasan pengguna yang tersedia di platform Tokopedia
- b. Analisis sentiment akan dilakukan pada ulasan pengguna dalam bahasa Indonesia.

- c. Penelitian ini akan menggunakan metode Naïve Bayes Classifier (NBC) sebagai pendekatan utama untuk analisis sentimen.
- d. Perancangan sistem dilakukan dengan menggunakan bahasa pemrograman python dan framework streamlit.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

- a. Melakukan analisis sentimen terhadap ulasan kepuasan pelanggan terhadap produk yang ditawarkan oleh toko Pengrajin.com di platform Tokopedia.
- b. Memperoleh pemahaman yang mendalam tentang persepsi pelanggan terhadap produk-produk Pengrajin.com berdasarkan analisis sentimen ulasan pelanggan.
- c. Menyediakan wawasan yang berguna bagi toko Pengrajin.com dalam meningkatkan kualitas produk, layanan, dan pengambilan keputusan bisnis berdasarkan umpan balik pelanggan.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Memberikan wawasan tentang tingkat kepuasan pelanggan terhadap produk Pengrajin.com di Tokopedia melalui analisis sentimen ulasan pelanggan.
- b. Membantu Pengrajin.com memahami persepsi pelanggan terhadap produk mereka dan mengidentifikasi area perbaikan yang diperlukan.
- c. Meningkatkan kualitas produk dan layanan Pengrajin.com dengan memanfaatkan informasi sentimen pelanggan.

1.6 Sistematika Penelitian

Penulisan Laporan tugas akhir yang terdiri dari beberapa sub-sub pembahasan dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Bab pertama pendahuluan, menguraikan tentang latar belakang masalah, rumusan masalah, ruang lingkup serta tujuan dan manfaat dalam penulisan penelitian Analisis sentiment review menggunakan algoritma Naïve Bayes Classifier.

BAB II KAJIAN HASIL PENELITIAN & LANDASAN TEORI

Bab ini membahas tentang teori-teori mengenai data mining, text mining dan analisis sentimen dengan metode *Naïve Bayes* dan beberapa metode lainnya. Yang berasal dari jurnal, buku, laporan terdahulu serta studi kepustakaan yang digunakan sebagai landasan teori dalam penelitian Analisis sentiment review menggunakan algoritma Naïve Bayes Classifier.

BAB III METODE PENELITIAN

Bab ini berisi tentang metodologi penelitian dan metodologi dalam menganalisa perhitungan yang digunakan dari penelitian Analisis sentiment review menggunakan algoritma Naïve Bayes Classifier.

BAB IV ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan analisis dan perancangan sistem yang berkaitan dan mendukung dalam pemecahan masalah terkait Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier : Studi Kasus Toko Pengrajin.Com.

BAB V IMPLEMENTASI DAN HASIL SERTA PEMBAHASAN

Bab ini memaparkan implementasi dan hasil serta pembahasan yang berkaitan dan mendukung dalam pemecahan masalah penelitian terkait Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier : Studi Kasus Toko Pengrajin.Com.

BAB VI KESIMPULAN

Pada bab ini berisi tentang kesimpulan dan saran sebagai pertimbangan untuk penelitian selanjutnya dari penelitian Analisis sentiment review menggunakan algoritma Naïve Bayes Classifier.

BAB II

KAJIAN HASIL PENELITIAN DAN LANDASAN TEORI

2.1 Kajian Hasil Penelitian

Penelitian sebelumnya atau kajian Pustaka merupakan hal yang sangat bermanfaat untuk menjadi perbandingan acuan yang memberikan gambaran terhadap hasil-hasil penelitian terdahulu menyangkut *e-commerce*, untuk melakukan penelitian perlu ada suatu bentuk hasil penelitian terdahulu yang dijadikan referensi pembanding, untuk itu pada bagian ini akan diberikan beberapa penelitian terdahulu yang berkaitan dengan rencana penelitian ini.

Penelitian ini dilakukan oleh Tito Dwiki Armawan (2022) dengan judul “Analisis Sentimen *Review* Pelanggan *E-COMMERCE* di Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*”. Penelitian ini berbicara tentang hal yang menjadi faktor kepuasan pelanggan terhadap suatu produk *e-commerce* dan juga mengetahui seberapa banyak pelanggan yang memberikan review positif atau negatif pada suatu produk tersebut. Penelitian ini bertujuan untuk membantu *e-commerce* melakukan analisis sentiment terhadap review pelanggan. Dengan menggunakan *website* id.priceprice.com dan play.google.com yang diambil mulai dari tahun 2017-2021 dan berdasarkan data yang diambil dari website di atas, dapat diartikan bahwa terdapat 5 (lima) *e-commerce* terpopuler pada saat ini yaitu *Tokopedia*, *Shopee*, *Bukalapak*, *Lazada*, dan *Blibli* yang akan digunakan penulis untuk melakukan penelitian kali ini. Pengelompokan review pelanggan menjadi sentiment negatif dan positif dilakukan dengan menerapkan metode *naïve bayes classifier* yang menghitung peluang berdasarkan frekuensi pengalaman sebelumnya. Penerapan *naïve bayes classifier* mendapatkan hasil tentang sentiment positif terbesar didapatkan oleh Lazada sebesar 97.0%, peringkat kedua bukalapak sebesar 94.6%, peringkat ketiga shopee sebesar 88.5%, peringkat keempat blibli.com sebesar 76.1%, peringkat kelima Tokopedia sebesar 34.4%, lalu hasil akurasi dari metode *naïve bayes* pada Lazada sebesar 56.23%, bukalapak sebesar 93.05%, shopee sebesar 87.82%, blibli.com sebesar 55.31%, dan

Tokopedia sebesar 94.94%. Hasil dari penelitian tersebut telah dievaluasi menggunakan *confusion matrix* dan divalidasi menggunakan *10-fold cross-validation*.

Penelitian berikutnya oleh Janah, dkk. (2020) dengan judul “Klasifikasi Emosi Ulasan Aplikasi Traveloka Pada Google Play Menggunakan *Naïve Bayes*” dalam Jurnal Informatika dan Sistem Informasi (JIFoSI) Vol. 1, No. 3. Penelitian ini menjelaskan bagaimana sistem dapat mendeteksi ulasan yang telah diberikan oleh pengguna terhadap aplikasi *Traveloka* pada *Google Play Store*. Penelitian ini bertujuan untuk membandingkan algoritma terbaik pada sentiment analysis ini menggunakan dua jenis kelas sentiment yaitu positif dan negatif. Dengan attribut data yang digunakan diantaranya nama, tanggal, rating dan ulasan dengan ulasan berbahasa Indonesia yang diambil, jumlah ulasan data aplikasi *Traveloka* yakni 2.662 ulasan dari bulan November 2019 hingga bulan Januari 2020. Pada penelitian ini, *text mining* yang dilakukan menggunakan metode *Naïve Bayes* yang digunakan ada 3 macam (Multinomial NB, Bernoulli NB, dan Gaussian NB), dengan tujuan untuk mendapatkan nilai akurasi yang terbaik dan memiliki sebaran nilai F1 score yang merata. Dari penelitian yang dilakukan, data ulasan aplikasi *Traveloka* yang didapat dari bulan November 2019 hingga bulan Januari 2020 yakni 2.662 ulasan, hasil akurasi terbaik didapatkan dengan menggunakan metode Multinomial NB dengan menggunakan metode *Hold Out* sebanyak 20% dan data test dengan akurasi sebesar 86% sedangkan metode Bernoulli NB akurasinya sebesar 69% dan Gaussian NB akurasinya sebesar 44%.

Penelitian ini dilakukan oleh Yulita, dkk. (2021) dengan judul “Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma *Naïve Bayes Classifier*” dalam **JDMSI**, Vol. 2, No. 2. Penelitian ini menjelaskan mengenai sudut pandang dan pendapat orang-orang di Indonesia mengenai vaksinasi COVID-19. Penelitian ini bertujuan untuk menganalisis pendapat masyarakat tentang vaksinasi COVID-19 di Indonesia. Analisis dilakukan terhadap data 3780 tweet yang berkaitan vaksinasi dengan menggunakan algoritma *Naïve Bayes Classifier*. Pengumpulan data pada penelitian ini dapat diproses dengan melakukan

pengambilan data set dari Kaggle. Data set tersebut merupakan tweet dari masyarakat dan tweet tersebut akan dikelompokkan dengan label positif, negatif dan netral menggunakan metode *Naïve Bayes Classifier*. Berdasarkan analisis, dapat diamati bahwa Sebagian besar tweet memiliki sikap positif (60,3 %), sementara jumlah tweet yang netral (33,4 %) melebihi jumlah *Tweet* yang menentang (5,4 %). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).

Penelitian selanjutnya dilakukan oleh Alzami, dkk. (2020) dengan judul “Sentiment Analysis Untuk Deteksi Ujaran Kebencian Pada Domain Politik” dalam Science And Engineering National Seminar 5 (SENS 5). Penelitian ini menjelaskan tentang bagaimana tulisan seseorang yang ada di *Twitter* dapat diidentifikasi dan dilakukan klasifikasi apakah tulisan tersebut termasuk kedalam ujaran kebencian atau tidak. Tujuan penelitian ini digunakan sebagai batu pijakan dalam filter ujaran kebencian sehingga para pengguna dapat menggunakan aplikasi media social dengan lebih nyaman. Penelitian ini, menggunakan *Sastrawi* untuk proses *stemming* tulisan di *Twitter*, kemudian melalui proses pembersihan karakter, *unigram* dan *naïve bayes* digunakan untuk tahap klasifikasi. Dari penelitian yang dilakukan, menghasilkan performa dengan recall sebesar 84.8%, precision sebesar 85.4% dan akurasi sebesar 85%. Dengan Performa yang cukup menggembirakan, dapat disimpulkan bahwa kombinasi *Sastrawi*, pembersihan karakter, *unigram* dan *naïve bayes* dapat digunakan untuk mendeteksi ujaran kebencian pada domain politik.

Penelitian ini dilakukan oleh Darwis, dkk. (2021) dengan judul “Penerapan Algoritma *Naïve Bayes* untuk Analisis Sentimen Review Data *Twitter* BMKG Nasional” dalam Jurnal TEKNO KOMPAK, Vol. 15, No. 1. Penelitian ini berbicara tentang penilaian atau opini masyarakat yang diambil dari tulisan di twitter mengenai pelayanan yang diberikan BMKG Nasional. Tujuan dari penelitian dengan Sentiment Analysis ini adalah untuk melakukan pencarian komentar negatif, positif dan netral dengan review data *Twitter* BMKG Nasional menggunakan algoritma *Naïve Bayes*, serta dapat menghasilkan klasifikasi berdasarkan nilai yang didapat dengan algoritma tersebut.

Metode yang digunakan dalam penelitian ini adalah klasifikasi data menggunakan algoritma Naïve Bayes Classifier (NBC). Sistem yang dikembangkan dengan menggunakan data internal yang diambil dari internet/Twitter untuk proses penentuan kalimat termasuk opini positif, netral atau negatif. Penentuan tersebut digolongkan sebagai proses pengklasifikasian. Serta menggunakan Application Python 3.74. Hasil penelitian ini masuk kedalam fine grained sentiment analysis yaitu analisis pada suatu kalimat komentar. Data tersebut akan diproses menggunakan text mining, kemudian dilanjutkan dengan mengklasifikasikan Tweet ke dalam tiga kelas, yaitu positif, negatif dan netral. Dari penelitian yang dilakukan dapat memberikan kemudahan bagi pengguna untuk melihat opini positif, negatif dan netral. Hasil uji akurasi pada metode naïve bayes untuk klasifikasi yaitu 69.97%.

Tabel 2. 1. Sumber Pustaka Primer

No	Judul	Penulis	Metode	Hasil/ Kesimpulan
1	Analisis Sentimen <i>Review</i> Pelanggan <i>E-COMMERC E</i> di Indonesia Menggunakan Algoritma <i>Naïve Bayes Classifier</i>	Tito Dwiki Harmawan (2022)	<i>Naïve Bayes Classifier</i>	Penerapan <i>naïve bayes classifier</i> mendapatkan hasil tentang sentiment positif terbesar didapatkan oleh Lazada sebesar 97.0%, peringkat kedua bukalapak sebesar 94.6%, peringkat ketiga shopee sebesar 88.5%, peringkat keempat blibli.com sebesar 76.1%, peringkat kelima Tokopedia sebesar 34.4%, lalu hasil akurasi dari metode <i>naïve bayes</i> pada Lazada sebesar 56.23%, bukalapak sebesar 93.05%, shopee sebesar 87.82%, blibli.com sebesar 55.31%, dan Tokopedia sebesar 94.94%.
2	Klasifikasi Emosi Ulasan Aplikasi Traveloka Pada Google Play Menggunakan <i>Naïve Bayes</i>	Ana Kholifatul Janah, Eka Dyar Wahyuni, dan Amalia Anjani Arifiyanti (2020)	<i>Naïve Bayes</i> (Multinomial NB, Bernoulli NB, dan Gaussian NB)	data ulasan aplikasi <i>Traveloka</i> yang didapat dari bulan November 2019 hingga bulan Januari 2020 yakni 2.662 ulasan, hasil akurasi terbaik didapatkan dengan menggunakan metode Multinomial NB dengan menggunakan metode <i>Hold Out</i> sebanyak 20% dan data test dengan akurasi sebesar 86% sedangkan metode Bernoulli NB akurasinya sebesar 69% dan Gaussian NB akurasinya sebesar 44%.

3	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma <i>Naïve Bayes Classifier</i>	Winda Yulita, Eko Dwi Nugroho dan Muhammad Habib Algifari (2021)	<i>Naïve Bayes Classifier</i>	Berdasarkan analisis, dapat diamati bahwa Sebagian besar tweet memiliki sikap positif (60,3 %), sementara jumlah tweet yang netral (33,4 %) melebihi jumlah <i>Tweet</i> yang menentang (5,4 %). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).
4	Sentiment Analysis Untuk Deteksi Ujaran Kebencian Pada Domain Politik	Farrikh Alzami, Nuanza Purinsyira P, Ricardus Anggi P, Rama Aria Megantara dan Dwi Puji Prabowo (2020)	<i>Naïve Bayes Classifier</i>	Dari penelitian yang dilakukan, menghasilkan performa dengan recall sebesar 84.8%, precision sebesar 85.4% dan akurasi sebesar 85%. Dengan Performa yang cukup menggembirakan, dapat disimpulkan bahwa kombinasi <i>Sastrawi</i> , pembersihan karakter, <i>unigram</i> dan <i>naïve bayes</i> dapat digunakan untuk mendeteksi ujaran kebencian pada domain politik.
5	Penerapan Algoritma <i>Naïve Bayes</i> untuk Analisis Sentimen Review Data <i>Twitter</i>	Dedi Darwis, Nery Siskawati dan Zaenal Abidin (2021)	<i>Naïve Bayes Classifier</i>	Dari penelitian yang dilakukan dapat memberikan kemudahan bagi pengguna untuk melihat opini positif, negatif dan netral. Hasil uji akurasi pada metode <i>naïve bayes</i> untuk klasifikasi yaitu 69.97%.

6	Yang Diusulkan: Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma <i>Naïve Bayes Classifier</i> : Studi Kasus Toko Pengrajin.Com (2023)	Muhammad Aulia (2023)	<i>Naïve Bayes Classifier</i>	-
---	--	--------------------------	---------------------------------------	---

2.2 Landasan Teori

2.2.1 E-Commerce

Menurut Dedi Riswandi (2019) *E-Commerce* adalah rangkaian Dinamika ekonomi, aplikasi dan proses bisnis menggabungkan perusahaan, dari konsumen dan masyarakat transaksi elektronik dan perdagangan barang, layanan/pelayanan dan informasi yang diberikan elektronik. Menurut Elfanso & Monica (2023), *E-commerce* adalah model bisnis yang memungkinkan perusahaan dan individu untuk membeli dan menjual barang melalui internet. Toko online beroperasi di empat segmen pasar utama. Bisnis ke bisnis, bisnis ke konsumen, konsumen ke konsumen, konsumen ke bisnis.

Menurut Kehista, dkk (2023), *E-commerce* adalah bisnis online menggunakan internet. Transaksi jual-beli dilakukan secara elektronik melalui platform online. Melalui e-commerce, pelaku bisnis dapat menawarkan produk dan layanan kepada konsumen secara global. Perkembangan teknologi mempengaruhi e-commerce, seperti pembayaran digital dan kecerdasan buatan. E-commerce juga membangun hubungan dengan konsumen melalui media sosial dan konten pemasaran. Ini adalah sektor bisnis yang berkembang pesat di era digital.

E-commerce (electronic commerce) merupakan proses yang memungkinkan teknologi-teknologi berbasis situs internet yang memfasilitasi perniagaan atau perdagangan. *E-commerce* memfasilitasi penggunaan dan implementasi proses baru bisnis. Hal ini mencakup pelaksanaan bisnis secara elektronik melintasi spektrum hubungan-hubungan antar perusahaan-perusahaan.

Menurut Salsabela & Solikin (2022), keberadaan online shop memiliki sejumlah kelebihan yang tak dapat diabaikan. Salah satunya adalah kemudahan dan keekonomisan akses yang diberikan kepada para pembeli, karena mereka hanya perlu melihat model barang melalui internet untuk memenuhi kebutuhan belanja mereka. Selain itu, dengan adanya online shop, konsumen juga dapat mengakses berbagai informasi detail tentang produk yang dijual, melihat ulasan dari pembeli lain, dan membandingkan harga dari berbagai penjual, semuanya dengan beberapa kali klik saja. Kelebihan-kelebihan tersebut menjadikan pengalaman belanja secara online lebih praktis, efisien, dan berbiaya lebih rendah dibandingkan dengan cara tradisional yang melibatkan kunjungan langsung ke toko fisik.

2.2.2 Tokopedia

Tokopedia adalah salah satu platform e-commerce terbesar di Indonesia yang memungkinkan penjual dan pembeli untuk bertransaksi secara online. Melalui Tokopedia, penjual dapat menjual berbagai produk mulai dari pakaian, elektronik, hingga kebutuhan rumah tangga, sementara pembeli dapat dengan mudah mencari dan membeli produk yang mereka butuhkan. Platform ini menawarkan fitur-fitur seperti sistem pencarian yang canggih, metode pembayaran yang aman dan beragam, serta dukungan yang responsif. Selain itu, Tokopedia juga aktif dalam mendukung pertumbuhan UMKM dengan menyediakan program pelatihan, dukungan pemasaran, dan solusi pengiriman yang terjangkau. Melalui upaya ini, Tokopedia telah berperan penting dalam memperluas pasar online di Indonesia dan memberikan peluang bisnis yang lebih besar bagi penjual lokal.

2.2.3 Pengrajin.Com

Pengrajin.com Official adalah sebuah toko di platform Tokopedia yang secara khusus menjual furnitur untuk rumah dan kantor. Toko ini menyediakan beragam pilihan furnitur berkualitas tinggi untuk memenuhi kebutuhan dekorasi dan fungsionalitas ruangan. Dari furnitur ruang tamu, kamar tidur, meja kerja, rak buku, hingga perabotan dapur, Pengrajin.com Official menawarkan produk-produk yang dirancang dengan baik dan terbuat dari bahan yang tahan lama. Toko Pengrajin.com Official di Tokopedia menawarkan berbagai macam gaya dan desain furnitur untuk memenuhi selera dan kebutuhan pelanggan. Mulai dari furnitur dengan desain klasik, minimalis, modern, hingga kontemporer, pelanggan dapat menemukan pilihan yang sesuai dengan gaya dan tema dekorasi ruangan mereka. Selain itu, produk-produk yang dijual oleh Pengrajin.com Official memiliki reputasi baik dalam hal kualitas dan kepuasan pelanggan.

2.2.4 Scraping

Scraping data adalah teknik otomatisasi pengambilan informasi dari website dengan menggunakan program komputer. Dengan proses ini, data dapat diekstraksi dari halaman web untuk digunakan dalam berbagai tujuan seperti penelitian, analisis pasar, pemantauan harga, dan pengumpulan berita. Namun, penting untuk mematuhi aturan etika dan hukum yang berlaku dalam penggunaan scraping data, termasuk menghormati hak cipta, privasi, dan kebijakan situs web yang dikunjungi.

2.2.5 Analisis Sentimen

Menurut Setiyawati & Cahyono (2023), Analisis sentimen adalah teknik penambangan data yang mencari dan mengekstrak data dari teks untuk memahami emosi sosial. Analisis sentimen juga dilakukan untuk lebih memahami sikap, pendapat, dan pendekatan teks informatif.

Menurut Sasmita, dkk (2022), Analisis sentimen adalah jenis penelitian atau studi yang melihat pada text mining. Ini adalah perhitungan matematis dari dua hal, yaitu pendapat dan perasaan yang terkandung dalam teks. Perhitungan matematisnya adalah analisis sentimen.

Analisis sentimen adalah jenis penelitian atau studi yang melihat pada text mining. Ini adalah perhitungan matematis dari dua hal, yaitu pendapat dan perasaan yang terkandung dalam teks. Perhitungan matematisnya adalah analisis sentiment oleh Romadloni, dkk (2019).

2.2.6 Naïve Bayes Classifier

Naïve Bayes merupakan metode untuk klasifikasi text dengan kecepatan pemrosesan yang tinggi jika dalam data besar. (Fitriyyah, Safriadi, & Pratama, 2019).

Menurut Sinaga, Sawaluddin dan Suwilo (2020) *Naïve Bayes* adalah metode yang digunakan untuk prediksi karena mengandung probabilistic sederhana yang diterapkan pada teorema bayes dengan ketergantungan yang kuat.

Pada penelitan ini, jenis metode naïve bayes yang digunakan yaitu *Gaussian NB*.

1. *Gaussian NB*

Adalah pengklasifikasi *Naïve Bayes* paling sederhana yang memiliki asumsi bahwa data dari masing-masing label diambil dari distribusi *Gaussian* sederhana. Metode *Naïve Bayes* ini dapat dipahami dengan menggunakan nama *conditional probabilities* (probabilitas bersyarat) yang dirumuskan sebagai berikut:

$$P(X) = \frac{P(H)P(H)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

P = Probabilitas

H = Hipotesis

X = Bukti

$P(H|X)$ = Probabilitas bahwa hipotesis H benar untuk bukti X

$P(X|H)$ = Probabilitas bahwa bukti X benar untuk hipotesis H

$P(H)$ = Probabilitas *prior* hipotesis H

$P(X)$ = Probabilitas *prior* bukti X

2.2.7 Text Preprocessing

Text Processing adalah suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur lagi dengan melalui serangkaian tahapan yang meliputi tahapan diantaranya :

- Case folding : merupakan proses yang dilakukan untuk menyeragamkan karakter pada data. Proses ini mengubah seluruh huruf menjadi huruf kecil.
- Tokenizing : merupakan tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.
- Filtering : merupakan tahap mengambil kata-kata penting dari hasil term.
- Stemming : adalah metode untuk mencari kata dasar dari sebuah kata. Proses ini dilakukan dengan cara menghilangkan imbuhan yang terdapat pada kata.

Text processing merupakan salah satu implementasi dari text mining. Text mining sendiri adalah suatu kegiatan menambang data, dimana data yang biasanya diambil berupa text yang bersumber dari dokumen-dokumen yang memiliki goals untuk mencari kata kunci yang mewakili dari sekumpulan document tersebut sehingga nantinya dapat dilakukan Analisa hubungan antara dokumen-dokumen tersebut.

2.2.8 TF-IDF

Term Weighting TF-IDF adalah salah satu pembobotan yang sering digunakan dan merupakan gabungan dari Term Frequency dan Inverse Document Frequency. TF-IDF terdiri dari frekuensi term dan inverse dokumen yang didapatkan dari membagi

seluruh jumlah dokumen terhadap jumlah dokumen yang memiliki term tersebut. Dalam TI-IDF bobot akan ditemukan dalam persamaan berikut:

$$W_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right) \dots \dots \dots (2)$$

Keterangan:

$\{tf\}_{i,j}$ = Bobot dari istilah i dalam dokumen j

$\{df\}_i$ = Frekuensi munculnya istilah i dalam dokumen j

N = Total istilah pada dokumen

2.2.9 Natural Language Processing (NLP)

Natural Language Processing adalah satu cabang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada pemrosesan bahasa. NLP digunakan untuk memudahkan interaksi dan komunikasi antara mesin dan manusia. Berkat NLP, komunikasi antara manusia dengan mesin menjadi lebih natural sehingga manusia merasa sedang tidak berkomunikasi dengan mesin.

Secara umum, ada tahapan dari analisis sintaksis yang sering digunakan, yaitu :

- *Stemming*
- *Lemmatization*
- *Tokenization*
- *Parsing*

2.2.10 Confusion Matrix

Menurut Holmes, Agius, dan Cromptvoets (2020), *confusion matrix* dan indeks yang berasal darinya adalah alat statistic untuk analisis pasangan pengamatan. Ketika tujuannya adalah untuk membandingkan dua data rahasia (dengan proses yang berbeda, berbeda operator, waktu yang berbeda, atau sesuatu yang serupa). *Confusion matrix*,

atau bisa disebut juga matriks kesalahan, adalah table kemungkinan, yang merupakan alat statistic untuk analisis pengamatan berpasangan. Matriks kebingungan diusulkan dan didefinisikan sebagai ukuran kualitas standar untuk data spasial. Pada Tabel 2 merupakan *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda.

Tabel 2. 2. Confusion Matrix

TP (<i>True Positive</i>)	FP (<i>False Positive</i>) <i>Type I Error</i>
FN (<i>False Negative</i>) <i>Type II Error</i>	TN (<i>True Negative</i>)

Keterangan :

True Positive (TP): Merupakan data positif yang diprediksi benar.

True Negative (TN): Merupakan data negatif yang diprediksi benar.

False Positive (FP) – *Type I Error* : Merupakan data negative namun diprediksi sebagai data positif.

False Negative (FN) – *Type II Error* : Merupakan data positif namun diprediksi sebagai data negatif.

2.2.11 Akurasi

Akurasi digunakan untuk mengetahui nilai akurasi dari hasil klasifikasi pada penelitian. Berikut adalah rumus dari akurasi *confusion matrix*

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(3)$$

Keterangan :

True Positive (TP) : Jumlah data bernilai benar yang memang memiliki nilai kebenaran datanya adalah benar.

False Negative (FN) : jumlah data salah yang memang memiliki nilai kebenaran datanya adalah salah.

False Positive (FP) : Jumlah data benar yang dianggap sistem memiliki nilai kebenaran datanya adalah benar.

True Negative (TN) : jumlah data salah yang dianggap sistem memiliki nilai kebenaran benar.

(TP + TN) : Jumlah prediksi yang hasilnya benar.

(TP + TN + FP + FN) : Total keseluruhan jumlah prediksi.

Untuk mengukur baik atau tidaknya algoritma atau teknik klasifikasi dapat dilihat dari akurasi prediksi yang dilakukan oleh algoritma tersebut. Semakin tinggi nilai akurasinya maka semakin baik algoritma yang dijalankan. Jika terdapat dua akurasi dan akurasi yang pertama sebesar 50% dan akurasi yang kedua 70%. Maka akan dipilih akurasi yang kedua karena memiliki akurasi yang paling besar. Sedangkan akurasi pertama sebesar 50% maka itu seperti peluangnya dalam melakukan prediksi benar makin kecil dan peluang prediksi benar hanya sebesar 50%.

2.2.12 Python

Menurut Onno W. Purbo (2019) *Python* adalah *general-purpose, high-level programming language*. Filosofi desain python menitikberatkan pada *code readability*, dan syntax yang memungkinkan *programmer* untuk mengekspresikan konsepnya dengan *lines of code* yang lebih sedikit daripada bahasa lainnya. Python dikatakan sebagai bahasa yang menggabungkan kemampuan, keahlian dan sintaks kode yang

sangat jelas, serta dilengkapi dengan fungsi library standar yang besar dan lengkap. Adapun library yang digunakan dalam penelitian ini, diantaranya :

- *Matplotlib.pyplot*

Matplotlib adalah library *Python* yang berfokus pada visualisasi data seperti plot grafik. Library ini dirancang untuk menghasilkan plot grafik yang sesuai pada publikasi jurnal atau artikel ilmiah. Visualisasi dalam *matplotlib* adalah sebuah grafik yang terdapat pada sebuah satu sumbu atau lebih. Setiap sumbu memiliki sumbu horizontal (x) dan sumbu vertical (y), dan yang direpresentasikan menjadi warna dan glyphs seperti marker atau line dan polygon.

Matplotlib dapat digunakan untuk melakukan visualisasi data secara 2D maupun 3D dan menghasilkan suatu gambar yang berkualitas bahkan dapat disimpan dengan format gambar seperti JPEG, JPG, dan PNG. Beberapa jenis visualisasi di *Python* menggunakan *Matplotlib* diantaranya line plot, scatter plot, bar plot, box plot dan histogram.

- *Pandas*

Pandas adalah paket *Python* open source yang digunakan untuk menganalisis data serta membangun sebuah machine learning. *Pandas* dibuat berdasarkan satu package lain bernama Numpy, yang mendukung arrays multi dimensi. Cara kerja *Pandas* yaitu merubah data berupa CSV, TSV, atau SQL ke obyek *Python* dengan rows dan column yang disebut sebagai data frame. Objek ini terlihat sangat mirip dengan table yang terdapat pada software statistika seperti Excel atau SPSS.

Program library *Pandas* dapat dijalankan menggunakan berbagai text editor, namun sangat disarankan untuk menggunakan Jupyter Notebook. Hal ini dikarenakan kemampuan Jupyter untuk meng-execute kode dari satu cell tertentu tanpa perlu menjalankan file secara keseluruhan. Library *Pandas*

memiliki beberapa fungsi diantaranya data cleansing, data fill, normalisasi data, Penggabungan dan penyatuan data, visualisasi data serta analisis statistik.

- *Sklearn*

Sklearn merupakan sebuah module bahasa pemrograman *Python* yang dibangun berdasarkan NumPy, SciPy, dan Matplotlib. Modul ini digunakan untuk melakukan processing data ataupun melakukan training data untuk kebutuhan machine learning atau data science.

Sklearn difokuskan pada machine learning, misalnya pemodelan data. Ini tidak melihat bagaimana proses pemuatan, penanganan, manipulasi, dan visualisasi data. Adapun *sklearn* lebih berfokus pada algoritma machine learning diantaranya regresi, classification, decision tree, neural networks, svm dan naïve bayes.

- *Wordcloud*

Wordcloud merupakan representasi visual dari kata teks yang biasanya dimanfaatkan untuk menggambarkan metadata kumpulan kata kunci atau *keywords* atau *tags* pada sebuah dokumen, untuk memberikan visualisasi bentuk teks secara bebas

- *Numpy*

Numpy adalah sebuah library pada *Python* yang berfungsi untuk melakukan operasi vector dan matriks dengan mengolah array dan array multidimensi serta melakukan proses komputasi numerik. Library ini dapat memudahkan operasi komputasi pada data dan melakukan akses secara acak, dan elemen array merupakan sebuah nilai yang independent sehingga penyimpanannya dianggap sangat efisien.

- *Selenium*

Selenium Python adalah modul Python yang digunakan untuk otomatisasi pengujian aplikasi web. Dengan Selenium Python, pengguna dapat mengontrol browser, melakukan interaksi dengan elemen halaman, dan memverifikasi hasil

pengujian. Kombinasi Selenium dan Python memungkinkan pengguna untuk mengotomatisasi pengujian aplikasi web dengan efisien dan meningkatkan akurasi pengujian.

- *Streamlit*

Streamlit Python adalah framework open-source untuk membuat antarmuka pengguna interaktif dalam pengembangan aplikasi data. Dengan Streamlit, pengembang dapat dengan mudah membuat aplikasi web interaktif untuk visualisasi dan eksplorasi data dengan cepat. Streamlit menyediakan API Python yang sederhana dan intuitif, serta fitur-fitur seperti grafik interaktif, widget kontrol, dan penerbitan aplikasi secara online. Keunggulan Streamlit adalah kemudahan penggunaannya dalam membuat aplikasi web interaktif tanpa pengetahuan mendalam tentang pengembangan web.

- *Imblearn*

Imblearn Python adalah pustaka yang digunakan untuk menangani ketidakseimbangan kelas dalam dataset. Dengan menggunakan imblearn, pengguna dapat melakukan over-sampling untuk meningkatkan jumlah sampel kelas minoritas dan under-sampling untuk mengurangi jumlah sampel kelas mayoritas. Teknik over-sampling menciptakan sampel sintetis, sedangkan under-sampling menghapus sampel acak. Tujuannya adalah mencapai keseimbangan antara kelas minoritas dan mayoritas. Dengan menggunakan teknik-teknik ini, imblearn membantu meningkatkan kinerja model pembelajaran mesin dalam memprediksi kelas minoritas dengan lebih akurat.

- *BeautifulSoup*

BeautifulSoup adalah pustaka Python yang digunakan untuk melakukan web scraping. Dengan BeautifulSoup, pengguna dapat mengekstraksi informasi terstruktur dari kode HTML atau XML halaman web. Pustaka ini dapat mengatasi markup yang tidak teratur dan menyediakan berbagai fungsi untuk

mencari, mengambil, dan memanipulasi data. BeautifulSoup memiliki integrasi yang baik dengan pustaka lain dalam ekosistem Python.

- *Sastrawi*

Sastrawi Python adalah pustaka Python yang digunakan untuk pemrosesan dan analisis teks dalam bahasa Indonesia. Pustaka ini menyediakan fungsi untuk normalisasi teks, stemming, dan tokenisasi. Sastrawi Python mempermudah pengguna dalam mengimplementasikan pemrosesan teks bahasa Indonesia dengan akurasi dan efisiensi.

- *NLTK*

NLTK Python adalah pustaka yang digunakan untuk pemrosesan bahasa alami dalam Python. NLTK menyediakan alat dan sumber daya untuk tugas seperti tokenisasi, stemming, analisis sintaksis, analisis sentimen, dan pembentukan model bahasa. Pustaka ini juga mendukung koleksi korpus dan teknik pemodelan bahasa. NLTK memungkinkan pengembang membangun aplikasi pemrosesan teks yang lebih kompleks dengan mudah.

2.2.13 Visualisasi

Visualisasi data adalah proses penyajian data dalam bentuk grafik yang membuat informasi mudah dimengerti, hal ini membantu menjelaskan tentang fakta dan menentukan arah tindakan. Definisi visualisasi data menjelaskan tentang pentingnya data dengan menempatkan data dalam konteks visual. Hal ini melibatkan penciptaan dan studi representasi visual dari data yang dikenal sebagai informasi. Visualisasi data memungkinkan pengguna untuk memperoleh pengetahuan yang lebih banyak mengenai data mentah yang edidapatkan dari berbagai sumber. Visualisasi dapat dilakukan dengan menggunakan *dashboard*, di mana teks, pola, dan korelasi yang tidak terdeteksi dapat dengan mudah divisualisasikan dengan menggunakan perangkat lunak visualisasi. Visualisasi juga merupakan bentuk upaya manusia dalam

mendeskripsikan maksud tertentu menjadi sebuah bentuk informasi yang mudah dipahami.

Visualisasi digunakan oleh penulis untuk mempermudah menjelaskan data yang akan dihasilkan. Agar para *e-commerce* dapat melihat dengan jelas dan tidak perlu banyak penjelasan untuk tau informasi atau data apa yang akan ditunjukkan.

BAB III METODE PENELITIAN

3.1 Bahan/Data

3.1.1 Data yang Diperoleh

Penelitian ini menggunakan data ulasan toko Pengrajin.Com yang diambil melalui platform Tokopedia sebagai sumber data primer. Data ulasan pelanggan yang dikumpulkan dari toko Pengrajin.Com di Tokopedia akan digunakan sebagai dataset utama dalam penelitian ini. Dengan menggunakan data ulasan yang diperoleh secara langsung dari toko Pengrajin.Com, penelitian ini dapat memberikan wawasan yang lebih spesifik dan relevan terkait dengan pengalaman pelanggan dalam berinteraksi dengan produk-produk yang ditawarkan oleh toko tersebut. Contoh data dapat dilihat pada gambar di bawah.

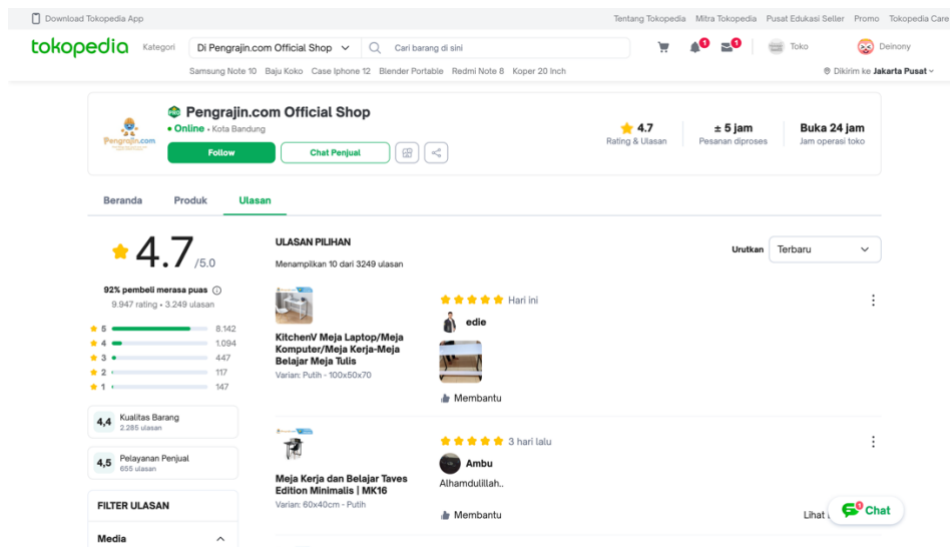
Nama Pelanggan	Produk	↓ Ulasan	Rating
Nandar	Bantal / Alas Duduk Lesehan Panjang Sofa/Kursi rebounded FC07	yup... sesuai orderan. seller responsif 🙌🙌🙌	5
Ari	Tissue Box Kotak Tempat Penyimpanan Tisu Kayu minimalist besar	Ya. Gitu lah. Keren	5
Hendri	Meja tamu / kopi kayu minimalis oval edition MW02	Ya ada uang ada barang. Kualitas? tidak diragukan kualitasnya	1
M***a	Kaki meja besi / metal table leg heavy duty editions TL17	yg buat skrup nya ga lurus	3
t***i	Kaki meja hairpin Minimalis Heavy Duty Craftsman	yak suip.. barang telah sampai dengan selamat, terima kasih	5
M	Meja belajar tulis dan Laptop kayu minimalis - MSB11	ya worth it lah	5
a***a	Meja kerja / belajar / komputer / laptop minimalis MSB11	ya so so lah, cuman mungkin ketebalan meja agak di tambah	4
fadjar	Kaki meja hairpin Minimalis Heavy Duty Craftsman	x kedua rekom	5
Ari	Wooden Hook / Gantungan Kayu Serbaguna Minimalis Edition HA13	Wuih. Bikin Ok dekorasi.	5
N***a	Alas Bantal Duduk Lesehan Kotak Sofa/Kursi rebounded FC03	Wow! Produknya sesuai foto & deskripsi. Jahitannya rapi. Se	5
Dinan	Meja Hairpin Coffee Table Minimalis MK27	Worth it dengan harganya dari segi kualitasnya	5

Gambar 3. 1 Data Ulasan Tokopedia

3.1.2 Prosedur Pengumpulan Data

Penelitian ini menggunakan metode scraping menggunakan Selenium untuk mengumpulkan data ulasan toko Pengrajin.Com di situs web Tokopedia. Metode ini memungkinkan pengendalian browser untuk mengakses halaman toko Pengrajin.Com,

mengambil ulasan pelanggan, dan mengumpulkan data secara otomatis. Dengan demikian, penelitian ini dapat menganalisis ulasan pelanggan dengan efisien dan mendalam untuk memperoleh wawasan yang relevan terkait dengan pengalaman pelanggan di toko Pengrajin.Com di Tokopedia. Tampilan Ketika melakukan scraping data dapat dilihat pada gambar di bawah.



Gambar 3. 2 Ulasan Toko Pengrajin.Com

Selanjutnya, Selenium akan berinteraksi dengan halaman toko Pengrajin.Com di situs web Tokopedia untuk mencari ulasan yang relevan. Proses ini mungkin melibatkan mengklik pada halaman produk yang tepat, menavigasi melalui halaman ulasan pelanggan, dan mengumpulkan data ulasan yang ada. Selenium akan mengekstrak teks ulasan dari elemen HTML yang sesuai dan menyimpannya dalam format csv untuk preprocessing lebih lanjut. Langkah-langkah tambahan seperti pembersihan data, penghapusan karakter khusus, atau penggabungan ulasan yang terpisah juga dapat dilakukan untuk memastikan data siap untuk analisis sentimen.

3.2 Aturan Bisnis (Bussiness Rule)

3.2.1 Analisis Sistem Saat ini

Analisis sistem saat ini terhadap aturan bisnis untuk studi kasus ulasan toko Pengrajin.com di Tokopedia menunjukkan beberapa hal yang perlu diperhatikan :

1. Responsifitas terhadap ulasan: Sistem perlu ditingkatkan agar lebih responsif dalam menanggapi ulasan pelanggan.
2. Kualitas analisis sentimen: Sistem perlu meningkatkan tingkat akurasi analisis sentimen ulasan.
3. Pemrosesan skala besar: Sistem harus mampu mengelola jumlah ulasan yang besar secara efisien.

Dalam mengoptimalkan sistem, perlu dilakukan pembaruan pada aspek-aspek tersebut untuk meningkatkan kualitas analisis sentimen, memberikan respon yang lebih baik terhadap ulasan pelanggan, dan memperbaiki pengalaman pelanggan secara keseluruhan.

3.2.2 Kelemahan/Kekurangan Sistem Saat ini

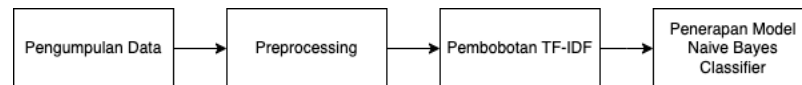
Sistem analisis sentimen ulasan toko Pengrajin.com di platform Tokopedia memiliki beberapa kelemahan yang perlu diperhatikan. Poin-poin penting yang perlu diperhatikan adalah:

1. Keterbatasan skala: Sistem mungkin tidak mampu mengelola jumlah ulasan yang besar secara efisien.
2. Keterbatasan bahasa: Sistem saat ini mungkin hanya mendukung analisis sentimen pada ulasan dalam bahasa Indonesia.
3. Ketidaktepatan dalam analisis: Sistem menggunakan algoritma Naïve Bayes Classifier yang mungkin tidak selalu mengenali konteks dan nuansa tertentu dalam ulasan dengan akurat.

Dalam pengembangan sistem, penting untuk mengatasi kelemahan-kelemahan ini agar analisis sentimen ulasan dapat menjadi lebih akurat dan komprehensif.

3.3 Tahapan Penelitian

Tahapan penelitian adalah proses yang dilakukan selama penelitian yang diawali dengan pengumpulan data dengan melakukan scraping hingga penerapan model naïve bayes classifier. Adapun urutan tahapan penelitian dapat dilihat pada gambar di bawah.



Gambar 3. 3 Tahapan Penelitian

Pertama, data ulasan pelanggan toko Pengrajin.com dikumpulkan melalui platform Tokopedia dengan menggunakan teknik scraping menggunakan Selenium. Setelah itu, data ulasan tersebut menjalani tahap preprocessing, di mana dilakukan pembersihan data seperti penghilangan karakter khusus, tautan, dan kata-kata yang tidak relevan dalam bahasa Indonesia. Tahap selanjutnya adalah pembobotan TF-IDF, di mana setiap kata dalam ulasan diberi bobot berdasarkan frekuensi kemunculannya dan kepentingannya dalam seluruh dataset. Terakhir, diterapkan model Naive Bayes Classifier untuk mengklasifikasikan ulasan menjadi kategori sentimen positif atau negatif. Dengan tahapan-tahapan tersebut, penelitian ini dapat memberikan wawasan tentang persepsi pelanggan terhadap toko Pengrajin.com di Tokopedia dan membantu pengambilan keputusan bisnis yang lebih baik.

BAB IV

ANALISIS DAN PERANCANGAN SISTEM

4.1 Analisis Sistem

4.1.1 Analisis Sistem yang Berjalan

Analisis sistem saat ini terhadap aturan bisnis untuk studi kasus ulasan toko Pengrajin.com di Tokopedia menunjukkan beberapa hal yang perlu diperhatikan :

1. Responsifitas terhadap ulasan: Sistem perlu ditingkatkan agar lebih responsif dalam menanggapi ulasan pelanggan.
2. Kualitas analisis sentimen: Sistem perlu meningkatkan tingkat akurasi analisis sentimen ulasan.
3. Pemrosesan skala besar: Sistem harus mampu mengelola jumlah ulasan yang besar secara efisien.

Dalam mengoptimalkan sistem, perlu dilakukan pembaruan pada aspek-aspek tersebut untuk meningkatkan kualitas analisis sentimen, memberikan respon yang lebih baik terhadap ulasan pelanggan, dan memperbaiki pengalaman pelanggan secara keseluruhan.

4.1.2 Analisis Sistem yang Diusulkan

Sistem yang diusulkan adalah sebuah sistem sentiment analysis dengan menggunakan algoritma machine learning yaitu metode Naïve Bayes Classifier untuk menganalisis sentiment dari data ulasan produk Pengrajin.com pada Tokopedia. Hal ini diharapkan dapat menghasilkan analisis sentiment yang baik dari opini para pelanggan terhadap suatu produk. Sistem yang dibuat dilakukan dengan tahapan-tahapan secara langsung yaitu dari pelabelan, kemudian dilakukan preprocessing pada data tersebut sebelum dilakukan penerapan model Naïve Bayes Classifier.

a. Analisis Fungsional

Kebutuhan fungsional dapat dijelaskan dalam 3 hal yaitu kebutuhan jenis masukan, proses-proses yang dibutuhkan, dan luaran yang diharapkan. Adapun kebutuhan tersebut diantaranya :

1. Kebutuhan Masukan

Data ini digunakan untuk mencatat persediaan produk yang ditawarkan oleh toko Pengrajin.com di platform Tokopedia dan juga digunakan dalam mencatat transaksi yang terkait dengan produk tersebut. Data tersebut meliputi informasi mengenai jumlah stok produk yang tersedia, perubahan persediaan akibat transaksi penjualan atau pengiriman barang, dan data terkait lainnya yang diperlukan untuk memantau dan mengelola persediaan toko.

2. Kebutuhan Proses

Pada penelitian ini, sistem melakukan proses analisis sentimen terhadap ulasan pelanggan pada toko Pengrajin.com di Tokopedia. Tujuannya adalah untuk menentukan apakah pesan ulasan tersebut memiliki nada emosional yang positif atau negatif. Dengan menganalisis teks ulasan secara digital, sistem dapat memberikan informasi tentang sentimen pelanggan terhadap produk yang ditawarkan oleh toko Pengrajin.com. Hal ini membantu toko dalam memahami persepsi pelanggan, mengidentifikasi kekuatan dan kelemahan produk, serta meningkatkan keputusan bisnis yang lebih baik untuk memenuhi kebutuhan pelanggan dengan lebih baik.

3. Kebutuhan Luaran

Pada tahapan ini, hasil dari pengolahan data ulasan toko Pengrajin.com di Tokopedia sudah dapat ditentukan apakah suatu ulasan produk bersifat positif atau negatif. Dengan menerapkan metode analisis sentimen, sistem dapat mengevaluasi dan mengklasifikasikan ulasan pelanggan berdasarkan polaritasnya. Hal ini memungkinkan toko Pengrajin.com untuk memahami apakah ulasan pelanggan mengandung sentimen positif atau negatif terhadap produk yang mereka tawarkan. Dengan mengetahui hasil analisis sentimen,

toko dapat merespon ulasan pelanggan dengan lebih baik, mengambil tindakan yang diperlukan untuk meningkatkan kualitas produk, serta mengambil keputusan bisnis yang lebih tepat guna.

b. Analisis Non-Fungsional

1. Kebutuhan Perangkat Lunak

Perangkat lunak merupakan alat penunjang untuk membangun atau melakukan analisis pada sebuah data, perangkat lunak yang digunakan dalam project kali ini ialah :

- **Google Chrome**

Google chrome adalah sebuah *web browser* yang berguna untuk menjelajah dunia maya. Karena pembuat dan pengembang google chrome adalah perusahaan Google, maka dipastikan mempunyai kualitas yang sangat baik.

- **Python 3.7**

Python merupakan bahasa *interpreter* yang artinya kode *python* akan dieksekusi beris perbaris. Bila terdapat kesalahan program di tengah eksekusi, maka program akan berhenti, kecuali menggunakan penanganan eksepsi.

- **Visual Studio Code**

Visual Studio Code (VS Code) adalah editor teks sumber terbuka yang dikembangkan oleh Microsoft. Ini ringan, cepat, dan mendukung banyak bahasa pemrograman. Fitur-fiturnya meliputi penyorotan sintaks, pemeriksaan kesalahan, penyelesaian kode otomatis, integrasi Git, ekstensi, dan tema. VS Code populer di kalangan pengembang karena produktivitas yang tinggi dalam menulis dan mengedit kode.

- **Safari**

Safari adalah peramban web yang dikembangkan oleh Apple untuk perangkat macOS dan iOS. Ia menawarkan antarmuka pengguna yang intuitif, kinerja cepat, keamanan tinggi, serta integrasi yang baik dengan ekosistem Apple.

2. Kebutuhan Perangkat Keras

Perangkat keras yaitu peralatan komputer yang secara fisik yang berfungsi untuk melakukan analisis pada sebuah data. Sistem perangkat keras terdiri dari bagian input, bagian proses, bagian output dan bagian untuk melakukan percobaan aplikasi. Perangkat keras yang dibutuhkan ialah :

- Laptop Macbook Pro (13 inci, 2017, Dua port thunderbolt 3).
- Intel Core i5 dual-core 2,3 GHz, Turbo Boost hingga 3,6 GHz, dengan eDRAM 64 MB.
- SSD 512 GB.
- RAM LPDDR3 2133 MHz 8gb.

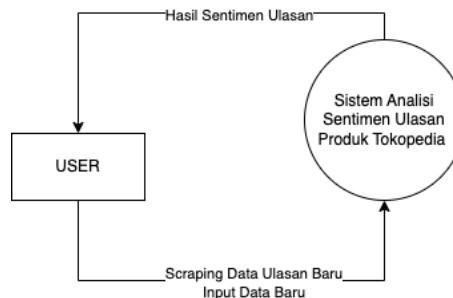
4.2 Desain Sistem

4.2.1 Perancangan Logik

Perancangan logika merupakan tahap dalam pengembangan sistem di mana alur logika sistem yang akan diimplementasikan dirancang. Pada tahap ini, digunakan metode perancangan seperti Data Flow Diagram (DFD) dan Flowchart. DFD digunakan untuk menggambarkan aliran data antara komponen sistem, sedangkan Flowchart digunakan untuk mewakili alur logika atau langkah-langkah dalam sistem. Dengan melakukan perancangan logika menggunakan metode ini, memungkinkan pengembang sistem untuk memiliki pemahaman yang jelas tentang bagaimana data mengalir dan bagaimana proses logika dijalankan dalam sistem yang dibangun.

- **Diagram Konteks**

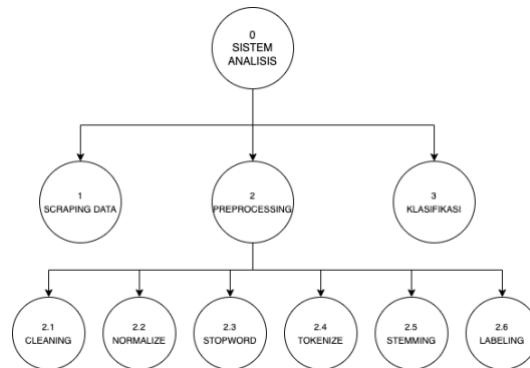
Diagram Konteks adalah rancangan sistem yang memberikan gambaran tentang entitas yang berinteraksi dengan sistem yang dibangun. Dalam konteks penelitian ini, Diagram Konteks menggambarkan entitas yang terlibat dalam melakukan analisis sentimen, yaitu pembuat sistem dan sistem yang dibangun. Diagram Konteks memberikan visualisasi tentang bagaimana interaksi antara entitas pembuat sistem dengan sistem yang melakukan analisis sentimen. Diagram konteks dapat dilihat pada gambar di bawah.



Gambar 4. 1 Diagram Konteks

- **Diagram Jenjang**

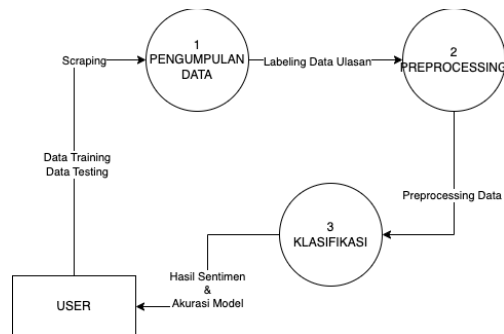
Diagram jenjang adalah sebuah representasi visual yang menunjukkan hubungan dan urutan tahap proses sistem secara berjenjang. Dalam konteks penelitian ini, Diagram jenjang akan menggambarkan tahapan-tahapan yang harus dilalui dalam sistem yang akan dibangun. Setiap tahapan akan diurutkan secara berjenjang, dimulai dari pengumpulan data, preprocessing, pembobotan TF-IDF, hingga penerapan model Naive Bayes Classifier. Diagram jenjang memberikan pemahaman yang lebih baik tentang bagaimana setiap tahapan saling terkait dan bergantung satu sama lain dalam proses analisis sentimen. Rancangan diagram jenjang dapat dilihat pada gambar di bawah.



Gambar 4. 2 Diagram Jenjang

- **Diagram Aliran Data Level 1**

Diagram Aliran Data (DFD) level 1 digunakan untuk menggambarkan hubungan data yang masuk dan keluar dari sistem serta proses yang terjadi di dalamnya. Dalam konteks penelitian ini, DFD level 1 menggambarkan hubungan antara entitas data input yang digunakan untuk melakukan pra-pemrosesan dan menghasilkan klasifikasi data menggunakan metode Naïve Bayes Classifier. DFD level 1 memberikan gambaran visual tentang bagaimana data mengalir dan diproses dalam sistem. DFD level 1 dapat dilihat pada gambar di bawah.

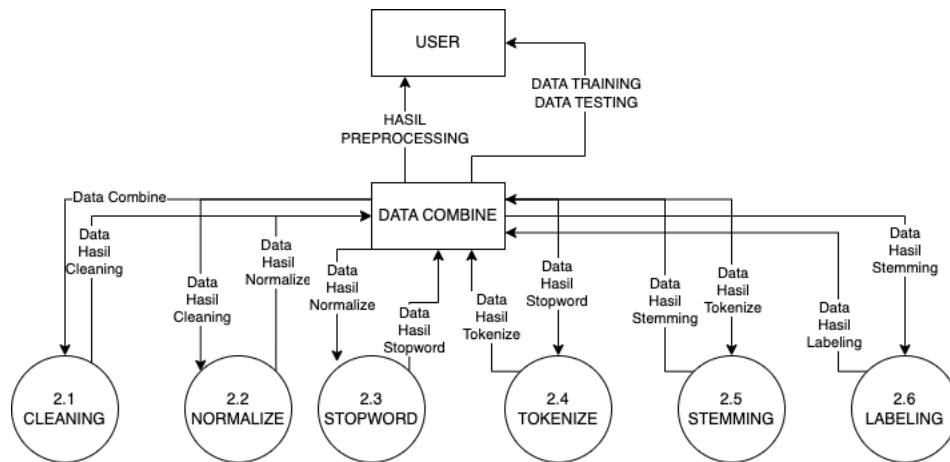


Gambar 4. 3 Diagram Aliran Data Level 1

- **Diagram Aliran Data Level 2 Proses 2**

Diagram Aliran Data (DFD) level 2, khususnya proses 2, memberikan gambaran lebih rinci tentang bagaimana data diproses dalam tahap preprocessing. Proses ini

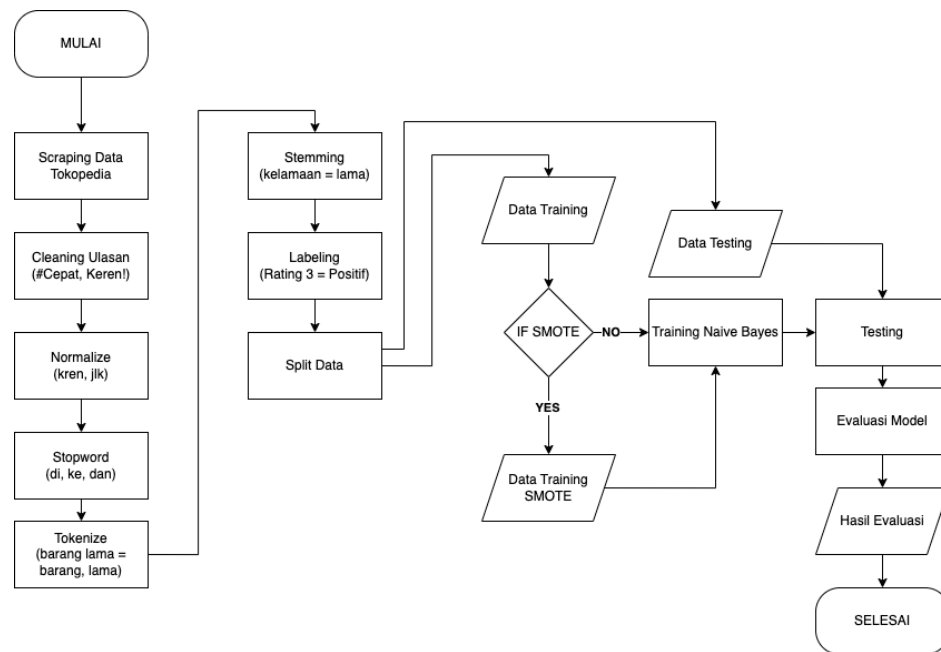
melibatkan pengolahan data yang telah diperoleh sehingga dapat dihasilkan data yang siap untuk dilakukan klasifikasi. Dengan menggunakan DFD level 2, kita dapat melihat alur data secara lebih terperinci pada tahap preprocessing. Diagram ini memberikan pemahaman tentang bagaimana data mengalir dan diolah dalam sistem sebelum dilakukan klasifikasi menggunakan metode Naïve Bayes Classifier. Diagram aliran data dapat dilihat pada gambar di bawah.



Gambar 4. 4 Diagram Aliran Data Level 2

• Flowchart Sistem Analisis Sentimen

Proses Analisis Sentimen dimulai dengan scraping data ulasan dari Tokopedia. Data ulasan tersebut kemudian melalui proses preprocessing dan pelabelan untuk persiapan analisis. Selanjutnya, dilakukan pembobotan menggunakan metode TF-IDF untuk menghitung frekuensi dan bobot kata-kata dalam ulasan. Pembobotan ini berguna dalam melakukan klasifikasi menggunakan metode Naïve Bayes Classifier. Setelah proses klasifikasi selesai, akan dihasilkan data yang telah dianalisis dengan tujuan untuk memperoleh informasi tentang sentimen ulasan pelanggan. Adapun gambaran sistem Flowchart dapat dilihat pada gambar di bawah.



Gambar 4. 5 Flowchart Sistem

BAB V

IMPLEMENTASI DAN HASIL SERTA PEMBAHASAN

5.1 Implementasi

Implementasi merupakan tahapan penting setelah perancangan aplikasi. Pada tahap ini, dilakukan implementasi bagian inti sistem menggunakan bahasa pemrograman Python. Dalam implementasi ini, potongan kode akan digunakan untuk menjelaskan langkah-langkah yang dilakukan dalam mengembangkan aplikasi tersebut.

5.1.1 Implementasi Proses Scraping

Pada tahap implementasi, dilakukan proses scraping data ulasan toko Pengrajin.com di Tokopedia menggunakan bahasa pemrograman Python dan library Selenium. Selenium digunakan untuk mengendalikan browser dan melakukan tugas-tugas seperti mengklik, memasukkan teks, dan menelusuri halaman web. Dengan menggunakan Selenium, sistem dapat mengakses halaman ulasan toko Pengrajin.com, menemukan elemen HTML yang berisi ulasan, dan mengekstrak teks ulasan tersebut. Selanjutnya, data ulasan yang berhasil diambil akan disimpan dalam format yang sesuai untuk proses analisis sentimen selanjutnya. Adapun tampilan *script* scraping dapat dilihat pada gambar di bawah.

```

96 if selected == 'Crawling':
97     col1, col2 = st.columns([1,8])
98     with col1:
99         st.image('img/tokopedia.png', width=80)
100     with col2:
101         st.title("Crawling Data")
102
103     # https://www.tokopedia.com/pengrajincom/review
104     url = st.text_input("URL")
105     tombol = st.button("Crawling")
106     if tombol :
107         options = webdriver.ChromeOptions()
108         options.add_argument("--start-maximized")
109         driver = webdriver.Chrome(options=options)
110         driver.get(url)
111
112         rating_mapping = {
113             'bintang 5': 5,
114             'bintang 4': 4,
115             'bintang 3': 3,
116             'bintang 2': 2,
117             'bintang 1': 1
118         }
119
120         data = []
121         for i in range(0, 200):
122             soup = BeautifulSoup(driver.page_source, "html.parser")
123             containers = soup.findAll('article', attrs={'class': 'css-ccpe8t'})
124
125             for container in containers:
126                 review_container = container.find('span', attrs={'data-testid': 'lblItemUlasan'})
127                 review = review_container.text.strip() if review_container else 'Tidak ada ulasan'
128
129                 nama_produk = container.find('p', attrs={'class': 'e1qvo2ff8'})
130                 produk = nama_produk.text.strip() if nama_produk else 'Produk tidak ditemukan'
131
132                 nama_pelanggan = container.find('span', attrs={'class': 'name'})
133                 pelanggan = nama_pelanggan.text.strip() if nama_pelanggan else 'Customer tidak ditemukan'
134
135                 rating_container = container.find('div', attrs={'data-testid': 'icnStarRating'})
136                 rating_label = rating_container['aria-label'] if rating_container else 'Tidak ada rating'
137                 rating = rating_mapping.get(rating_label, 'Tidak ada rating')
138
139                 data.append(
140                     (pelanggan, produk, review, rating)
141                 )
142
143                 time.sleep(2)
144                 driver.find_element(By.CSS_SELECTOR, "button[aria-label~='Laman berikutnya']").click()
145                 time.sleep(3)
146
147             df = pd.DataFrame(data, columns=['Nama Pelanggan', 'Produk', 'Ulasan', 'Rating'])
148             df.to_csv('data/Tokopedia.csv', index=False)
149             driver.close()
150
151             # Mengubah tipe data 'Rating' & memanggil atribut tertentu
152             showDf = pd.read_csv('data/Tokopedia.csv', dtype={'Rating': 'object'})
153             showDf = showDf[['Nama Pelanggan', 'Produk', 'Ulasan', 'Rating']]
154
155             # Menampilkan data
156             st.write(showDf)
157
158             # Pengkondisian Alert Crawling
159             jumlah_data = len(showDf)
160             if jumlah_data > 0:
161                 st.success(f"Crawling {jumlah_data} Baris Data Berhasil!")
162             else:
163                 st.warning("Crawling Data Gagal")

```

Gambar 5. 1 Script Scraping

5.1.2 Implementasi Proses Preprocessing

Pada tahap implementasi preprocessing, data ulasan toko Pengrajin.com yang telah berhasil diambil akan melalui serangkaian langkah untuk mempersiapkannya

sebelum dilakukan analisis sentimen. Proses preprocessing meliputi langkah-langkah seperti menghapus karakter khusus, mengubah teks menjadi huruf kecil, menghilangkan stopwords (kata-kata umum yang tidak memberikan makna signifikan), serta melakukan tokenisasi untuk memisahkan kata-kata dalam ulasan. Selain itu, langkah-langkah tambahan seperti stemming atau lemmatisasi dapat dilakukan untuk mengubah kata-kata menjadi bentuk dasar mereka. Semua langkah ini bertujuan untuk membersihkan dan mempersiapkan data ulasan agar siap digunakan dalam proses analisis sentimen selanjutnya dengan metode Naïve Bayes Classifier. Adapun tampilan *script* preprocessing dapat dilihat pada gambar di bawah.

```

35 # Download NLTK resources
36 nltk.download('punkt')
37 nltk.download('stopwords')
38
39 selected = option_menu(None, ["Crawling", "Dataset", "Preprocessing", "Machine Learning", "Prediction"],
40 icons=['cloud-upload', 'archive', 'gear', 'activity', 'kanban'],
41 menu_icon="cast", default_index=0, orientation="horizontal")
42
43 norm = { ' dgn ' : ' dengan ', ' seller ' : ' penjual ', ' service ' : ' pelayanan ', ' tp ' : ' tapi ', ' recommended ' : ' rekomendasi ',
44 ' blum ' : ' belum ', ' quality ' : ' kualitas ', ' bagus ' : ' bagus ', ' overall ' : ' akhirnya ', ' mantaaaaaappp ' : ' bagus ', ' joss ' : ' bagus
45 ' trima ' : ' terima kasih ', ' brg ' : ' barang ', ' ga ' : ' tidak ', ' dgn ' : ' dengan ', ' real ' : ' asli ', ' bnb ' : ' baru ',
46 ' recommended ' : ' rekomen ', ' recommend ' : ' rekomen ', ' good ' : ' bagus ',
47 ' eksis ' : ' ada ', ' becnilai ' : ' bernilai ', ' dg ' : ' dengan ', ' ori ' : ' asli ', ' setting ' : ' atur ', ' free ' : ' gratis ',
48 ' yg ' : ' yang ', ' tk ' : ' tempat ', ' awet ' : ' awet ', ' mantep ' : ' bagus ', ' mantapp ' : ' bagus ',
49 ' kl ' : ' kalo ', ' k ' : ' ke ', ' plg ' : ' pulang ', ' ajah ' : ' aja ', ' bgt ' : ' banget ', ' lah ' : ' lebih ', ' ayem ' : ' tenang ', ' dsana ' : ' disana ',
50 ' pas ' : ' saat ', ' bnb ' : ' baru ',
51 ' nggak ' : ' tidak ', ' karna ' : ' karena ', ' utk ' : ' untuk ',
52 ' dn ' : ' dan ', ' mlt ' : ' melihat ', ' pd ' : ' pada ', ' mndgr ' : ' mendengar ', ' crita ' : ' cerita ', ' dpt ' : ' dapat ', ' mksh ' : ' teri
53 ' sejh ' : ' sejarah ', ' mmh ' : ' memang ', ' sayapun ' : ' saya ', ' thn ' : ' tahun ', ' good ' : ' bagus ', ' awett ' : ' awet ',
54 ' halo ' : ' halo ', ' mntal ' : ' mental ', ' plus ' : ' dan ',
55 ' ayang ' : ' sayang ', ' Rekomended ' : ' direkomendasikan ', ' now ' : ' sekarang ', ' slalu ' : ' selalu ', ' photo ' : ' foto ', ' lah ' :
56 ' ya ' : ' saya ', ' slp ' : ' bagus ', ' bny ' : ' banyak ', ' tdk ' : ' tidak ', ' byk ' : ' banyak ',
57 ' pool ' : ' sekali ', ' pgn ' : ' ingin ', ' gue ' : ' saya ', ' bgmn ' : ' bagaimana ', ' ga ' : ' tidak ',
58 ' gak ' : ' tidak ', ' dr ' : ' dari ', ' yg ' : ' yang ', ' lu ' : ' kamu ', ' sya ' : ' saya ',
59 ' lancarr ' : ' lancar ', ' kayak ' : ' seperti ', ' ngawur ' : ' sembarangan ', ' k ' : ' ke ',
60 ' luasss ' : ' luas ', ' sy ' : ' saya ', ' thn ' : ' tahun ', ' males ' : ' malas ',
61 ' tgl ' : ' tanggal ', ' lg ' : ' lagi ', ' bgt ' : ' banget ', ' gua ' : ' saya ', ' ln ' : ' ', ' tpi ' : ' tapi ', ' standar ' : ' biasa ', ' sta
62
63 def normalisasi(text):
64     for i in norm:
65         text = text.replace(i, norm[i])
66     return text
67
68 def clean(text):
69     text = text.strip()
70     text = text.lower()
71     text = re.sub(r"[a-z-2]*", ' ', text)
72     return text
73
74 def labeling(rating):
75     if rating == '4' or rating == '5':
76         return 'Positif'
77     else:
78         return 'Negatif'
79
80
81
82
83 -def stopword(text):
84     stop_words = set(stopwords.words('indonesian'))
85     words = text.split()
86     filtered_words = [word for word in words if word.casefold() not in stop_words]
87     cleaned_text = ' '.join(filtered_words)
88     return cleaned_text
89
90 -def stemming(text):
91     stemmer = StemmerFactory().create_stemmer()
92     text = ' '.join(text)
93     stemmed_text = stemmer.stem(text)
94     return stemmed_text

```

```

173 - if selected == "Preprocessing":
174     st.title("Preprocessing Data")
175     df = pd.read_csv("data/tokopedia.csv", dtype={"Rating": "object"})
176     df = df[['Nama Pelanggan', 'Produk', 'Ulasan', 'Rating']]
177     st.write(df)
178
179     preprocessing = st.button("Preprocessing")
180     if preprocessing:
181         with st.spinner('Sedang melakukan preprocessing...'):
182             time.sleep(2)
183             # st.success("Preprocessing Berhasil & Data Dilaunch")
184             df['Ulasan'] = df['Ulasan'].fillna("")
185             df['Ulasan'] = df['Ulasan'].apply(clean)
186             st.write("")
187             st.write(f"----- CLEANNING -----")
188             st.write(df['Ulasan'])
189
190             df['Ulasan'] = df['Ulasan'].apply(normalisasi)
191             st.write("")
192             st.write(f"----- NORMALIZE -----")
193             st.write(df['Ulasan'])
194
195             df['Ulasan'] = df['Ulasan'].apply(stopword)
196             st.write("")
197             st.write(f"----- STOPWORD -----")
198             st.write(df['Ulasan'])
199
200             df['Ulasan'] = df['Ulasan'].apply(tokenisasi)
201             st.write("")
202             st.write(f"----- TOKENIZE -----")
203             st.write(df['Ulasan'])
204
205             # Melakukan stemming pada kolom "Ulasan"
206             df['Ulasan'] = df['Ulasan'].apply(stemming)
207             st.write("")
208             st.write(f"----- STEMMING -----")
209             st.write(df['Ulasan'])
210
211             df['Sentimen'] = df['Rating'].apply(labeling)
212             st.write("")
213             st.write(f"----- LABELING -----")
214             st.write(df[['Ulasan', 'Sentimen']])
215
216             # Menghentikan animasi loading
217             st.spinner(False)
218
219             # Menghentikan animasi loading
220             st.spinner(False)
221             df = df[['Ulasan', 'Sentimen']]
222             df.to_csv('data/Preprocessing.csv', index=False)
223
224             # Pengkondisian Alert Crawling
225             jumlah_data = len(df)
226             if jumlah_data > 0:
227                 st.success(f"Preprocessing {jumlah_data} Baris & Download Data Berhasil !")
228             else:
229                 st.warning("Preprocessing Data Gagal")
230
231 ---

```

Gambar 5. 2 Script Preprocessing

5.1.3 Implementasi Proses Machine Learning

Pada tahap implementasi proses Machine Learning, metode Naïve Bayes Classifier digunakan untuk melakukan analisis sentimen terhadap ulasan toko Pengrajin.com. Data ulasan diolah dan dibagi menjadi data pelatihan dan pengujian. Model Naïve Bayes Classifier dilatih menggunakan data pelatihan untuk mengklasifikasikan ulasan menjadi sentimen positif atau negatif. Kinerja model dievaluasi menggunakan metrik evaluasi. Hasil evaluasi memberikan informasi tentang akurasi dan kemampuan model dalam mengklasifikasikan ulasan. Dengan menggunakan metode ini, sistem dapat memberikan prediksi sentimen ulasan yang dapat digunakan untuk meningkatkan kualitas produk dan layanan toko Pengrajin.com. Adapun tampilan *script* Machine Learning dapat dilihat pada gambar di bawah.

```

228 if selected == 'Machine Learning':
229     df = pd.read_csv('data/Preprocessing.csv')
230     df = df.dropna()
231     # df = pd.concat([df[df['Sentimen'] == 'Negatif'].head(257), df[df['Sentimen'] == 'Positif'].head(400)])
232
233     # Menghitung jumlah baris data dengan sentimen positif dan negatif
234     num_positive = df[df['Sentimen'] == 'Positif'].shape[0]
235     num_negative = df[df['Sentimen'] == 'Negatif'].shape[0]
236
237     # Menyiapkan data untuk visualisasi
238     sentimen = ['Positif: {num_positive}', 'Negatif: {num_negative}']
239     jumlah_data = [num_positive, num_negative]
240
241     # Membuat bar chart dengan warna tiap label yang berbeda
242     colors = ['#3691d6', '#d63636']
243     plt.bar(sentimen, jumlah_data, color=colors)
244     plt.xlabel('Sentimen')
245     plt.ylabel('Jumlah Data')
246     plt.title('Perbandingan Label Sentimen')
247
248     # Menampilkan chart menggunakan Streamlit
249     st.pyplot(plt)
250
251     smote_option = st.selectbox('Pilih opsi untuk menggunakan SMOTE & IMBLEARNS atau tidak:', ('Tanpa SMOTE & IMBLEARNS', 'Dengan SMOTE & IMBLEARNS'))
252     tombol = st.button("Training Data")
253     if tombol:
254         with st.spinner("Sedang melatih model..."):
255             # Bagi data menjadi atribut dan label
256             X = df['Ulasan']
257             y = df['Sentimen']
258
259             # Encoding label kelas menjadi nilai biner
260             y = np.where(y == 'Positif', 1, 0)
261
262             # Inisialisasi CountVectorizer
263             cv = CountVectorizer()
264
265             # Transformasi teks menjadi representasi numerik menggunakan CountVectorizer
266             X_transformed = cv.fit_transform(X)
267
268             if smote_option == "Dengan SMOTE & IMBLEARNS":
269                 # Lakukan SMOTE untuk mengatasi ketidakseimbangan kelas
270                 # smote = SMOTE()
271                 over = RandomOverSampler(sampling_strategy=0.5, random_state=5)
272                 under = RandomUnderSampler(sampling_strategy=0.85, random_state=5)
273
274                 X_resampled, y_resampled = over.fit_resample(X_transformed, y)
275                 X_resampled, y_resampled = under.fit_resample(X_resampled, y_resampled)
276             else:
277                 X_resampled, y_resampled = X_transformed, y
278
279             # Bagi data menjadi data training dan data testing
280             X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2, stratify=y_resampled, random_state=42)
281
282             # Inisialisasi model Naive Bayes
283             nb_model = MultinomialNB()
284
285             # Latih model menggunakan data training
286             nb_model.fit(X_train.toarray(), y_train)
287
288             # Prediksi menggunakan data testing
289             y_pred = nb_model.predict(X_test.toarray())
290
291             # Hitung akurasi model
292             accuracy = accuracy_score(y_test, y_pred)
293
294             # Simpan model ke file
295             joblib.dump(nb_model, 'model.pkl')
296
297             # Tampilkan akurasi
298             st.success("Pelatihan Model Naive Bayes Selesai!")
299
300             # Visualisasi distribusi kelas sebelum dan sesudah SMOTE
301             fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(10, 5))
302
303             ax1.bar(np.unique(y), np.bincount(y), color=['#d63636', '#3691d6'])
304             ax1.set_xticks([0, 1])
305             ax1.set_xticklabels(['Negatif', 'Positif'])
306             ax1.set_title('Sebelum SMOTE')
307             ax1.set_xlabel('Kelas')
308             ax1.set_ylabel('Jumlah')
309
310             ax2.bar(np.unique(y_resampled), np.bincount(y_resampled), color=['#d63636', '#3691d6'])
311             ax2.set_xticks([0, 1])
312             ax2.set_xticklabels(['Negatif', 'Positif'])
313             ax2.set_title('Setelah SMOTE & IMBLEARNS' if smote_option == "Dengan SMOTE" else 'Tanpa SMOTE')
314             ax2.set_xlabel('Kelas')
315             ax2.set_ylabel('Jumlah')
316
317             st.pyplot(fig)

```

```

318 # Menghitung confusion matrix
319 cm = confusion_matrix(y_test, y_pred)
320
321 # Definisikan label kelas
322 labels = ['Negatif', 'Positif']
323
324 # Visualisasi confusion matrix menggunakan heatmap
325 plt.figure(figsize=(8, 6))
326 sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=labels, yticklabels=labels)
327 plt.xlabel('Predicted')
328 plt.ylabel('Actual')
329 plt.title('Confusion Matrix')
330 st.pyplot(plt)
331
332 # Tampilkan classification report
333 report = classification_report(y_test, y_pred, output_dict=True)
334 report_df = pd.DataFrame(report).transpose()
335
336 # Ubah classification report menjadi tabel
337 table = tabulate(report_df, headers='keys', tablefmt='psql')
338
339 # Tampilkan tabel classification report
340 st.write("Classification Report:")
341 st.code(table)

```

Gambar 5. 3 Script Machine Learning

5.1.4 Implementasi Proses Prediction

Pada tahap implementasi proses Prediction, model yang telah dilatih menggunakan metode Naïve Bayes Classifier digunakan untuk melakukan prediksi sentimen ulasan pada toko Pengrajin.com. Data ulasan baru yang belum pernah dilihat sebelumnya dimasukkan ke dalam model yang telah dilatih untuk mendapatkan prediksi sentimen positif atau negatif. Proses prediksi ini memanfaatkan pengetahuan yang telah diperoleh dari data pelatihan sebelumnya. Hasil prediksi dapat memberikan informasi tentang bagaimana ulasan pelanggan akan dinilai berdasarkan sentimen yang terkandung di dalamnya. Dengan demikian, toko Pengrajin.com dapat menggunakan hasil prediksi ini untuk memahami umpan balik pelanggan dan mengambil tindakan yang sesuai guna meningkatkan kualitas produk, layanan, dan pengambilan keputusan bisnis. Adapun tampilan *script* Prediction dapat dilihat pada gambar di bawah.

```

343 if selected == 'Prediction':
344     st.title("Prediksi Sentimen")
345     input_text = st.text_area("Masukkan Teks Ulasan")
346     df = pd.read_csv('data/Preprocessing.csv')
347     df = df.dropna()
348
349     if st.button("Prediksi"):
350         if input_text:
351             # Load model dari file
352             nb_model = joblib.load('model.pkl')
353
354             # Inisialisasi CountVectorizer dan latih dengan data latihan
355             cv = CountVectorizer()
356             X_transformed = cv.fit_transform(df['Ulasan'])
357
358             # Transformasi teks input menjadi representasi numerik menggunakan CountVectorizer yang sudah dilatih
359             input_transformed = cv.transform([input_text])
360
361             # Prediksi sentimen menggunakan model yang sudah dilatih
362             prediction = nb_model.predict(input_transformed.toarray())
363             sentimen_prediksi = 'Positif' if prediction[0] == 1 else 'Negatif'
364
365             # Menghitung probabilitas hasil prediksi sentimen
366             probabilities = nb_model.predict_proba(input_transformed.toarray())
367             proba_positif = probabilities[0, 1]
368             proba_negatif = probabilities[0, 0]
369
370             # Tampilkan hasil prediksi
371             st.success("Hasil Prediksi:")
372             # st.write("Sentimen:", sentimen_prediksi)
373             st.write("Positif:", round(proba_positif * 100, 2), "%")
374             st.write("Negatif:", round(proba_negatif * 100, 2), "%")
375
376             # Animasi presentasi nilai sentimen
377             fig, ax = plt.subplots()
378             ax.axis('off')
379
380             # Menentukan gambar yang akan ditampilkan berdasarkan sentimen prediksi
381             if sentimen_prediksi == 'Positif':
382                 img_path = 'img/smile.png'
383             else:
384                 img_path = 'img/angry.png'
385
386             img = plt.imread(img_path)
387             imagebox = OffsetImage(img, zoom=2) # Mengubah zoom menjadi 0.2 untuk ukuran yang lebih kecil
388
389             ab = AnnotationBox(imagebox, (0.5, 0.5), frameon=False)
390             ax.add_artist(ab)
391
392             def update(frame):
393                 ab.xybox = (frame / 100, 1)
394                 return ab,
395
396             ani = animation.FuncAnimation(fig, update, frames=50, interval=50, blit=True)
397
398             # Menampilkan animasi
399             st.pyplot(fig)
400             plt.close(fig)
401
402         else:
403             st.warning("Masukkan teks ulasan sebelum melakukan prediksi!")

```

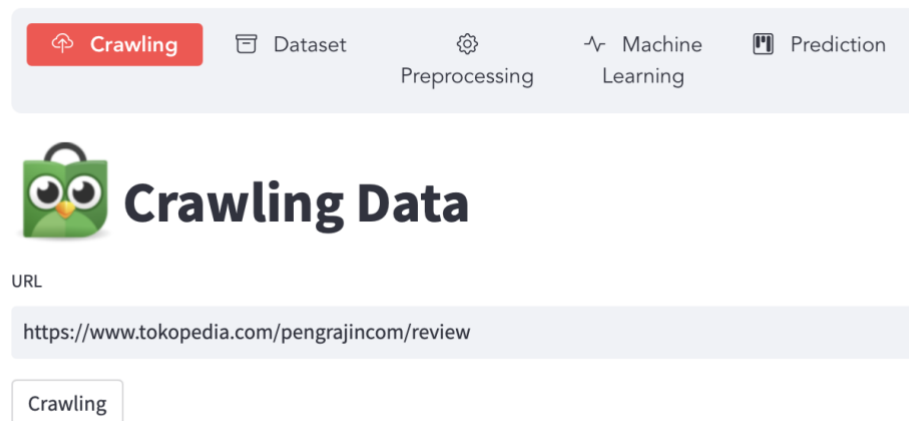
Gambar 5. 4 Script Prediction

5.2 Hasil

Sub bab hasil adalah tahap evaluasi setelah implementasi sistem, di mana dilakukan analisis terhadap output sistem untuk melihat kinerja dan akurasi dalam menganalisis sentimen ulasan toko Pengrajin.com. Evaluasi ini memberikan wawasan yang berguna bagi Pengrajin.com dalam pengambilan keputusan berdasarkan umpan balik pelanggan.

5.2.1 Implementasi Halaman Scraping

Implementasi halaman scraping melibatkan penggunaan fitur input URL, di mana pengguna diharuskan memasukkan URL halaman toko Pengrajin.Com sebelum melakukan proses crawling untuk mengumpulkan data pengguna. Proses crawling ini dilakukan untuk mengambil data penting seperti Nama Pelanggan, Produk, Ulasan, dan Rating. Melalui halaman scraping, pengguna dapat melihat tampilan yang menampilkan kolom input URL dan gambaran visual dari halaman scraping yang mencakup elemen-elemen penting yang akan diambil dari toko Pengrajin.Com, seperti yang ditunjukkan dalam gambar di bawah.

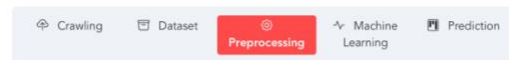


Gambar 5. 5 Tampilan Scraping

5.2.2 Implementasi Halaman Preprocessing

Implementasi halaman preprocessing adalah proses pengolahan data sesudah proses scraping. Tujuan dari preprocessing data adalah untuk mempersiapkan data agar memenuhi syarat sebelum dilakukan training ke model naïve bayes. Proses dari tahapan preprocessing diantaranya yaitu cleaning, normalize, stopword, tokenize,

stemming, dan labelling. Adapun tampilan dari halaman preprocessing dapat dilihat pada gambar di bawah.



Preprocessing Data

	Nama Pelanggan	Produk	Ulasan
56	noglaa	Gaming Table / Meja Gaming /Meja Kantor/ Meja Belajar/ meja komputer	kekokohan : 80/100 Presisi : 9
57	Ulung	Meja Kerja dan Belajar Hairpin Minimalis MK13	barang sesuai gambar, pema
58	****s	CUSTOM BANTAL / CUSTOM BUSA / CUSTOM SARUNG BANTAL ONLY	Barang sdh sampai dengan s
59	M***d	WITA - Kursi lesehan lipat sofa folding busa Dakron Premium edition	oke banget, seller ramah, em
60	a***i	Gaming Table / Meja Gaming /Meja Kantor/ Meja Belajar/ meja komputer	mantap
61	Halim	Rangka Kaki meja Minimalis Heavy Duty 1 set Craftman	lumayan ok, hollow sesuai dg
62	kenan	Bantal / Alas Duduk Lesehan Kotak Sofa/Kursi rebounded FCD6	bagus beud,, bahan bagus en
63	Fikry	Meja Belajar / Kerja / laptop Industrial Knockdown MK01	mejanya reyod, harus dilas la
64	Farikh	Kaki meja kayu / Wooden Table Legs square editions Heavy Duty TL04	Ketika dipasang di daun meji
65	****i	Rangka Kaki meja Minimalis Heavy Duty 1 set Craftman	koko kuat terpercaya

Preprocessing

CLEANING

	Ulasan
0	alhamdulillah
1	bagus
2	tidak ada ulasan
3	kaki nya kurang simetris mkn hrs lebih teliti lagi gan
4	bagus tebal
5	bagus cuma finishing ada benang hitam yang keluar tidak
6	respon seller bagus cepat tanggap produknya juga bagus
7	baguuuus
8	setelah dipasang tidak kokoh gampang goyang
9	ukuran pas sesuai penawaran warna sesuai permintaan sa

NORMALIZE

	Ulasan
0	alhamdulillah
1	bagus
2	tidak ada ulasan
3	kaki nya kurang simetris mkn hrs lebih teliti lagi gan
4	bagus tebal
5	bagus cuma finishing ada benang hitam yang keluar tidak ra
6	respon penjual bagus cepat tanggap produknya juga bagus
7	baguuuus
8	setelah dipasang tidak kokoh gampang goyang

9 ukuran saat sesuai penawaran warna sesuai permintaan cam

STOPWORD

Ulasan
0 alhamdulillah
1 bagus
2 ulasan
3 kaki nya simetris mkn hrs teliti gan
4 bagus tebal
5 bagus finishing benang hitam rapih
6 respon penjual bagus cepat tanggap
7 bagusss
8 dipasang kuku gampang gresang
9 ukuran sesuai penawaran warna ses

TOKENIZE

Ulasan
0 alhamdulillah
1 bagus
2 ulasan
3 kaki nya simetris mkn hrs teliti gan
4 bagus tebal
5 bagus finishing benang hitam rapih
6 respon penjual bagus cepat tanggap pri
7 bagusss
8 dipasang kuku gampang gresang
9 ukuran sesuai penawaran warna sesuai

STEMMING

Ulasan
0 alhamdulillah
1 bagus
2 ulas
3 kaki nya simetris mkn hrs teliti gan
4 bagus tebal
5 bagus finishing benang hitam rapih
6 respon jual bagus cepat tanggap pri
7 bagusss
8 pasang kuku gampang gresang
9 ukur sesuai tawar warna sesuai mint

LABELING

Ulasan	Sentimen
0 alhamdulillah	Positif
1 bagus	Positif
2 ulas	Positif
3 kaki nya simetris mkn hrs teliti gan	Negatif
4 bagus tebal	Positif
5 bagus finishing benang hitam rapih	Positif
6 respon jual bagus cepat tanggap pri	Positif
7 bagusss	Positif
8 pasang kuku gampang gresang	Negatif
9 ukur sesuai tawar warna sesuai mint	Positif

Preprocessing 1990 Baris & Download Data Berhasil!

Made with Streamlit

Gambar 5. 6 Tampilan Preprocessing

Pada tampilan diatas sistem melakukan preprocessing terhadap 1990 baris data. Data tersebut beberapa atribut sebelum dilakukan preprocessing, lalu hasil akhir dari preprocessing hanya mengambil setengah dari kondisi atribut awal yaitu mengambil

Ulasan dan Rating, karena peneliti ingin melakukan pengolahan teks pada atribut ulasan sebelum dilakukan pengujian sentiment.

5.2.3 Implementasi TF-IDF

TF (Term Frequency) adalah metode yang digunakan dalam analisis teks dan pemrosesan bahasa alami untuk mengukur sejauh mana kata-kata muncul dalam sebuah dokumen atau korpus teks. Pada dasarnya, TF menghitung frekuensi kemunculan kata tertentu dalam dokumen dan menggunakannya sebagai indikator penting untuk menentukan bobot atau relevansi kata-kata tersebut dalam konteks yang sedang dianalisis. TF menggambarkan seberapa sering suatu kata muncul dalam suatu teks relatif terhadap total kata-kata yang ada di teks tersebut. Umumnya, penghitungan TF dilakukan dengan menghitung jumlah kemunculan kata tertentu, lalu hasilnya dinormalisasi dengan membaginya dengan total kata dalam dokumen. Hal ini membantu mengurangi efek dari panjang dokumen yang berbeda-beda. Adapun tabel TF dapat dilihat pada gambar di bawah.

Tabel 5. 1 Hasil TF

Kata	TF
Bagus	692
Sesuai	500

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode yang digunakan dalam analisis teks dan pemrosesan bahasa alami untuk mengukur tingkat kepentingan suatu kata dalam suatu dokumen dalam konteks koleksi dokumen yang lebih besar. TF-IDF menggabungkan informasi tentang frekuensi kata dalam dokumen (TF) dan frekuensi kata dalam seluruh koleksi dokumen (IDF) untuk memberikan bobot atau skor yang mencerminkan tingkat keunikan kata dalam dokumen. Adapun hasil TF-IDF dapat dilihat pada tabel di bawah.

Tabel 5. 2 Hasil TF-IDF

Kata	TF-IDF
Bagus	162.8387
Sesuai	121.9255

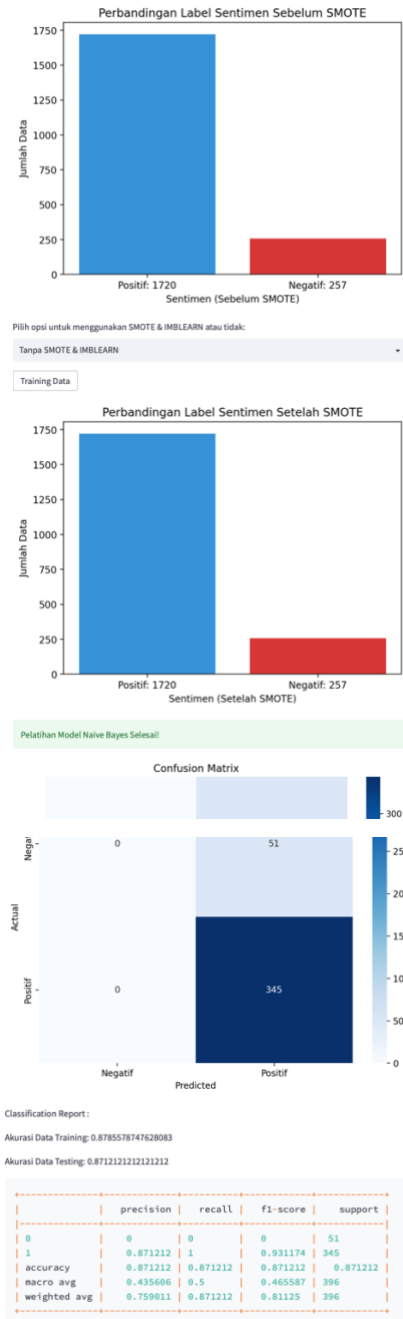
5.2.4 Training dan Testing

Setelah menyelesaikan tahapan preprocessing data, langkah selanjutnya adalah membagi data menjadi data pelatihan (training) dan data pengujian (testing) dengan rasio 80:20. Proses pemisahan data dilakukan menggunakan metode split, yang memungkinkan 80% data digunakan untuk melatih model dan 20% data digunakan untuk menguji kinerja model. Untuk memastikan reproduktibilitas, pengaturan random state 42 digunakan dalam proses pembagian data. Setelah pembagian data dilakukan, langkah berikutnya adalah melakukan pelatihan (fitting) data pengujian menggunakan model Naïve Bayes Classifier. Dengan demikian, model akan belajar dari data pelatihan dan kemudian diuji menggunakan data pengujian untuk mengukur kinerjanya dalam melakukan klasifikasi sentimen ulasan produk di Tokopedia.

5.2.5 Implementasi Halaman Machine Learning

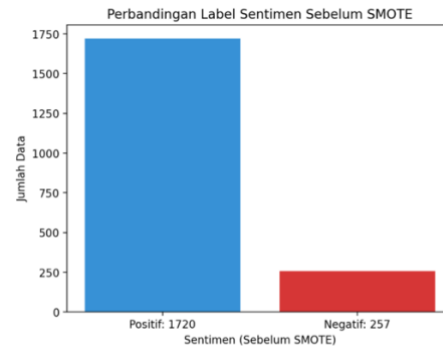
Implementasi halaman machine learning melibatkan langkah-langkah penting seperti pelatihan data yang telah melalui proses preprocessing sebelumnya. Halaman ini akan menampilkan visualisasi yang memberikan informasi mengenai jumlah data dengan sentimen positif dan negatif yang telah diproses. Selanjutnya, terdapat opsi untuk melakukan proses training dengan dua pilihan, yaitu menggunakan SMOTE & Imblearn atau tanpa menggunakan SMOTE & Imblearn. Melalui halaman machine learning, pengguna dapat melihat tampilan yang menggambarkan proses training yang sedang berlangsung dan mendapatkan hasil yang relevan. Halaman machine learning memberikan pengalaman interaktif yang memudahkan pengguna untuk menganalisis

sentimen ulasan produk di toko Pengrajin.Com. Adapun tampilan dari halam machine learning dapat dilihat pada gambar di bawah.



Gambar 5. 7 Tampilan Machine Learning tanpa SMOTE

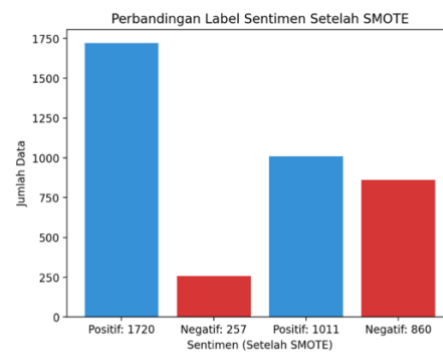
Pada gambar di atas dilakukan proses training data tanpa SMOTE & Imblearn dengan kondisi perbedaan data sentiment yang sangat jauh berbeda. Proses training pada tampilan di atas menghasilkan akurasi sebesar 87%.



Pilih opsi untuk menggunakan SMOTE & IMBLEARN atau tidak:

Dengan SMOTE & IMBLEARN

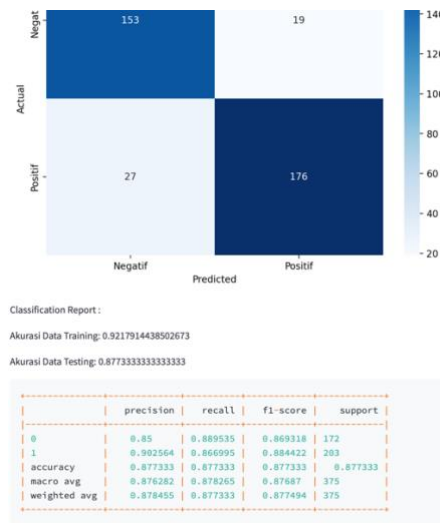
Training Data



Pelatihan Model Naive Bayes Selesai!

Confusion Matrix



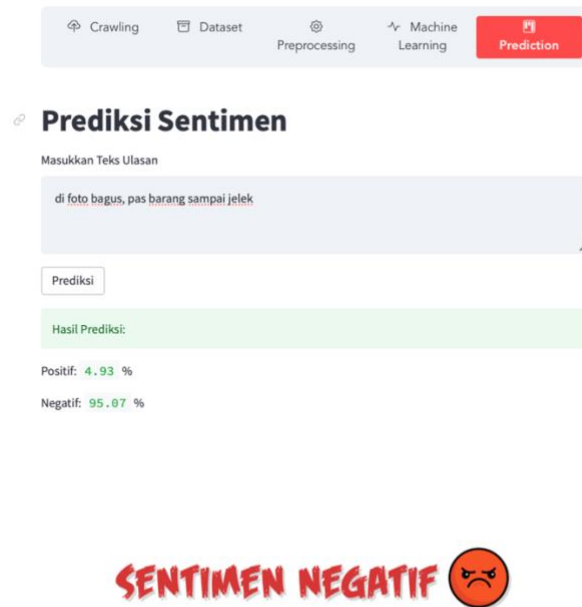


Gambar 5. 8 Tampilan Machine Learning dengan SMOTE

Pada gambar di atas dilakukan proses training data menggunakan SMOTE & Imblearn dengan kondisi jumlah sentiment data yang balance. Proses training pada tampilan di atas menghasilkan akurasi sebesar 87%.

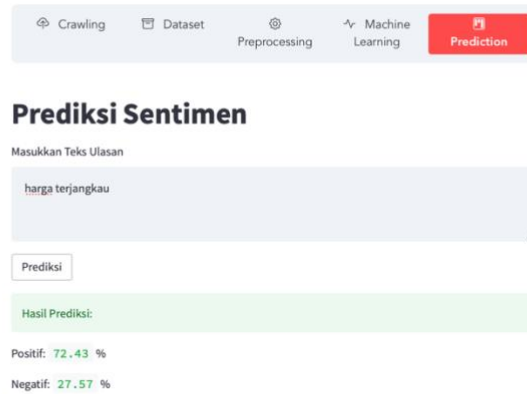
5.2.6 Implementasi Halaman Prediction

Sistem klasifikasi menggunakan algoritma Naïve Bayes ini diuji dengan menggunakan 298 data uji dan 1592 data latih. Akurasi yang dihasilkan dari pembagian split data tersebut adalah 86% tanpa menggunakan SMOTE & Imblearn, 88% dengan menggunakan SMOTE & Imblearn. Kemudian tahapan selanjutnya yaitu melakukan pengujian dan prediksi dari model yang sudah dibangun dengan cara melakukan input ulasan pada halaman prediction. Contoh pengujian oleh pengguna dapat dilihat pada gambar di bawah.



Gambar 5. 9 Tampilan Prediction Sentimen Negatif

Pada hasil pengujian di atas dengan input “di foto bagus, pas barang sampai jelek” menampilkan hasil prediksi Positif : 4.93% dan Negatif : 95.07%. Artinya sentiment negatif memiliki nilai yang lebih tinggi dan menghasilkan hasil *Sentimen Negatif*.



SENTIMEN POSITIF 😊

Gambar 5. 10 Tampilan Prediction Sentimen Positif

Pada hasil pengujian di atas dengan input “harga terjangkau” menampilkan hasil prediksi Positif : 72.43% dan Negatif : 27.57%. Artinya sentiment positif memiliki nilai yang lebih tinggi dan menghasilkan hasil *Sentimen Positif*.

BAB VI PENUTUP

6.1 Simpulan

Kesimpulan dari penelitian ini, yang berjudul "Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier: Studi Kasus Toko Pengrajin.com", adalah sebagai berikut:

1. Sistem yang dikembangkan mampu secara efektif mendeteksi sentimen positif dan negatif dari ulasan produk e-commerce menggunakan metode Naïve Bayes Classifier. Hal ini membuktikan kemampuan algoritma tersebut dalam melakukan analisis sentimen.
2. Dalam penelitian ini, Naïve Bayes Classifier berhasil mencapai tingkat keakuratan yang cukup tinggi, dengan hasil uji menggunakan data uji mencapai akurasi sebesar 87% tanpa menggunakan SMOTE & Imblearn, 87% dengan menggunakan SMOTE & Imblearn. Hal ini menunjukkan bahwa sistem analisis sentimen yang dikembangkan dapat memberikan hasil yang dapat diandalkan.
3. Sistem ini memiliki nilai akurasi, presisi, dan recall yang baik, sehingga dapat dijadikan sebagai komponen penting dalam sebuah sistem analisis sentimen. Kelebihan ini memberikan kepercayaan kepada pengguna sistem untuk mengandalkan analisis sentimen produk pada toko Pengrajin.com.

6.2 Saran

Adapun saran pada penelitian yang berjudul "Analisis Sentimen Ulasan Produk Tokopedia Menggunakan Algoritma Naïve Bayes Classifier: Studi Kasus Toko Pengrajin.com", adalah sebagai berikut:

1. Menerapkan metode klasifikasi yang lebih komprehensif: Selain menggunakan algoritma Naïve Bayes Classifier, disarankan untuk menjelajahi dan membandingkan kinerja dengan metode klasifikasi lainnya, seperti Support Vector

Machines (SVM), Random Forest, atau Neural Networks. Hal ini dapat memberikan gambaran yang lebih lengkap tentang kinerja sistem dalam menganalisis sentimen ulasan produk.

2. Memperluas cakupan data ulasan: Untuk memperoleh generalisasi yang lebih baik, disarankan untuk mengumpulkan data ulasan dari berbagai toko di Tokopedia, bukan hanya dari satu toko Pengrajin.com. Hal ini akan meningkatkan representativitas analisis sentimen terhadap produk-produk di platform yang lebih luas.
3. Melakukan analisis lebih mendalam terhadap hasil sentimen: Selain mengidentifikasi sentimen positif dan negatif, disarankan untuk melakukan analisis lebih mendalam terhadap hasil sentimen, seperti pengelompokan ulasan berdasarkan aspek tertentu (misalnya, kualitas produk, pelayanan pelanggan, atau harga). Hal ini dapat memberikan wawasan yang lebih spesifik dan bernilai tambah bagi pemilik toko dan pengguna platform

DAFTAR PUSTAKA

- Alzami, F., P, N. P., P, R. A., Megantara, R. A., & Prabowo, D. P. (2020). Sentiment Analysis untuk Deteksi Ujaran Kebencian pada Domain Politik. *Science And Engineering National Seminar 5*, 213-218.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal Tekno Kompak*, 131-145.
- Elfanso, E., & Monica, L. (2023). Analisis dampak transaksi E-Commerce terhadap penerimaan pajak pertambahan nilai. *JURNAL PROGRAM STUDI AKUNTANSI POLITEKNIK GANESHA*, 841-850.
- Harmawan, T. D. (2022). Analisis Sentiment Review Pelanggan E-Commerce di Indonesia Menggunakan Algoritma Naive Bayes Classifier.
- Janah, A. K. (2020). DETEKSI EMOSI PADA ULASAN APLIKASI ETICKETING PADA SITUS GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES (Studi kasus Aplikasi Traveloka). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 312-322.
- Kehista, A. P., Fauzi, A., & Tamara, A. (2023). Analisis Keamanan Data Pribadi pada Pengguna E-Commerce: Ancaman, Risiko, Strategi Kemanan (Literature Review). *JURNAL ILMU MANAJEMEN TERAPAN*, 625-632.
- Permadi, V. A. (2022). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. *Jurnal Buana Informatika*, 141-151.
- Qamal, M., Fuadi, W., & Muslia. (2021). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. *Jurnal Teknologi Terapan & Sains*, 124-132.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter. *Jurnal INVOTEK Polbeng*, 50-59.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, KNN Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *Jurnal IKRA-ITH Informatika*.
- Riswandi, D. (2019). Transaksi ON-LINE (E-Commerce) : Peluang Dan Tantangan Dalam Perspektif Ekonomi Islam. *Jurnal Econetica*.
- Salsabela, K., & Solikin. (2022). Sistem Informasi Penjualan Online Pada CV. Palugada Enterprise Jakarta Dengan Metode Extreme Programming. *Journal of Information Management*, 71-80.
- Setiyawati, D., & Cahyono, N. (2023). Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia. *Indonesian Journal of Computer Science*, 262-272.
- Sasmita, A. B., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan

Metode Naive Bayes. *JURNAL PENGEMBANGAN TEKNOLOGI INFORMASI DAN ILMU KOMPUTER*, 1208-1214.

Yulita, W., Nugroho, E. D., & Algifari, M. H. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *JDMSI*.