

LAPORAN SEMINAR TEMATIK/INFORMATIC CAPSTONE PROJECT

Semester Ganjil TA. 2022/2033

**ANALISIS SENTIMENT REVIEW PRODUCT ONLINE SHOP
MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER**

Dosen Pembimbing: Dr. Ir. Arief Hermawan. ST., MT., IPU



MUHAMMAD AULIA (5200411482)

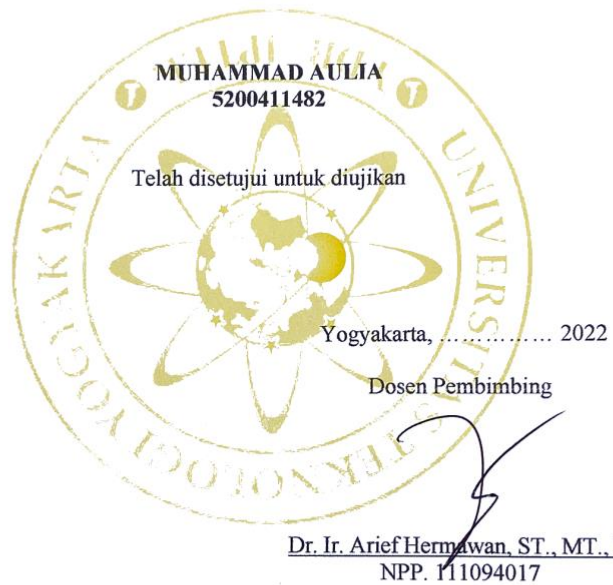
PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS TEKNOLOGI YOGYAKARTA
YOGYAKARTA

2022

**PROYEK INFORMATIKA
SEMESTER GANJIL 2022**

**ANALISIS SENTIMENT REVIEW PRODUCT ONLINE SHOP
MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER**

Diajukan Oleh



KATA PENGANTAR

Puji syukur dipanjatkan atas kehadiran Allah SWT, karena dengan limpahan karunia-Nya penulis dapat menyelesaikan Informatic Capstone Project dengan judul Analisis Sentiment Review Product Online Shop Menggunakan Algoritma Naïve Bayes Classifier.

Penyusunan Laporan Informatic Capstone Project 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.
- b. Bapak Endy Marlina, Dr., MT. selaku Dekan Fakultas Sains & Teknologi Universitas Teknologi Yogyakarta.
- c. Ibu Dr. Enny Itje Sela S.Si, M.Kom., selaku ketua Program Studi Informatika Fakultas Sains & Teknologi Universitas Teknologi Yogyakarta.
- d. Bapak Arief Hermawan, Dr., S.T., M.T. selaku Dosen Pembimbing yang telah berkenan membimbing dan mengarahkan penulis dalam penyelesaian Informatic Capstone Project.
- e. Kedua orang tua yang selalu support dan memberikan dukungan serta doa demi kelancaran Informatic Capstone Project.
- f. Semua keluarga yang selalu support dan memberikan dukungan doa demi kelancaran Informatic Capstone Project.
- g. Teman-teman yang telah memberikan dukungan serta memberikan semangat, motivasi dan juga membantu dalam penulisan serta penyusunan laporan Informatic Capstone Project.
- h. Seluruh pihak yang tidak dapat disebutkan satu persatu, yang telah mendukung dan memberikan masukan kepada penulis selama pengerjaan laporan Informatic Capstone Project.

Akhir kata, penulis menyadari bahwa sepenuhnya akan terbatasnya pengetahuan penyusunan sehingga tidak menutup kemungkinan, jika ada kesalahan serta kekurangan dalam penyusunan laporan Informatic Capstone Project, untuk itu sumbang saran dari pembaca sangat diharapkan sebagai bahan pelajaran berharga di masa yang akan datang.

Yogyakarta, 5 Januari 2023

Penulis

ABSTRAK

Saat ini masyarakat di Indonesia bahkan dunia lebih senang berbelanja barang-barang yang mereka inginkan menggunakan *online shop*. 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. Analisis sentiment atau opinion mining merupakan analisa terhadap suatu pendapat seseorang pada sesuatu hal tertentu yang dituangkan kedalam sebuah pendapat, perkataan atau komentar untuk memberikan masukan tersendiri bagi pihak yang sedang di berikan opini. Pada analisis sentiment terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan analisis sentimen perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui feedback masyarakat atau pun para ahli. Naïve Bayes adalah metode yang digunakan untuk prediksi karena mengandung probabilistic sederhana yang diterapkan pada teorema bayes dengan ketergantungan yang kuat. Hasil tes dari uji coba pada berbagai perubahan parameter menampilkan hasil akurasi tertinggi yaitu **84%**. Kemudian pada penelitian ini, algoritma *Naïve Bayes Classifier* terbukti merupakan algoritma yang akurat karena tingkat keakuratan sistem analisis sentimen produk e-commerce menggunakan algoritma naïve bayes yang diuji menggunakan data uji mendapatkan hasil akurasi sebesar **84%** dengan tingkat kebenaran 1000 untuk label positif dan 925 untuk label negatif.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
KATA PENGANTAR	iii
ABSTRAK	v
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup	2
1.4 Tujuan dan Manfaat Penelitian	2
1.5 Sistematika Penelitian	3
BAB II TINJAUAN PUSTAKA DAN TEORI	4
2.1 Tinjauan Pustaka	4
2.2 Teori	11
2.2.1 E-Commerce	11
2.2.2 Analisis Sentiment	12
2.2.3 Naïve Bayes Classifier	12
2.2.4 Text Preprocessing	13
2.2.5 TF-IDF	14
2.2.6 <i>Natural Language Processing (NLP)</i>	14
2.2.7 <i>Confusion Matrix</i>	15
2.2.8 Akurasi	16
2.2.9 <i>Google Colab</i>	16
2.2.10 <i>Python</i>	17
2.2.11 <i>Visualisasi</i>	19
BAB III METODE PENELITIAN	20
3.1 Kerangka Penelitian	20
3.2 Data Penelitian	20
3.2.1 Sumber Data	20
3.2.2 Cara Mendapatkan Data	21
3.2.3 Waktu Pengumpulan Data	21
3.3 Arsitektur Model	21
3.4 Analisis dan Perancangan	22
3.4.1 Kebutuhan Fungsional	22
3.4.2 Kebutuhan Non-Fungsional	22
3.4.3 Perancangan Konseptual	23
BAB IV PRODUK APLIKASI	28
4.1 Hasil	28
4.1.1 Tampilan Awal Prototipe	28
4.1.2 Proses Pengumpulan Data	29
4.1.3 Labeling	29
4.1.4 Preprocessing Teks	30
4.1.5 TF-IDF	33

4.1.6	Visualisasi	34
4.1.7	Pembuatan Model	37
4.1.8	Pelatihan Model (<i>Naïve Bayes Classifier</i>)	37
4.1.9	Prototype Pengujian Model.....	39
4.2	Pembahasan Hasil	40
4.3	Pengembangan ke Tugas Akhir	41
BAB V SIMPULAN		42

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Penelitian	20
Gambar 3. 2 Arsitektur Model	21
Gambar 3. 3 Preprocessing Data	25
Gambar 3. 4 TF-IDF	26
Gambar 3. 5 Klasifikasi Data	27
Gambar 4. 1 Labeling	29
Gambar 4. 2 Lower Teks	30
Gambar 4. 3 Normalisasi	31
Gambar 4. 4 Stopwords	31
Gambar 4. 5 Tokenize	32
Gambar 4. 6 Hasil Tokenize	32
Gambar 4. 7 Stemming	33
Gambar 4. 8 Menghitung Kata	33
Gambar 4. 9 Hasil Perhitungan Kata	34
Gambar 4. 10 Proses TF	34
Gambar 4. 11 Hasil TF	34
Gambar 4. 12 Visualisasi Negatif	35
Gambar 4. 13 Hasil Visualisasi Negatif	35
Gambar 4. 14 Visualisasi Positif	35
Gambar 4. 15 Hasil Visualisasi Positif	36
Gambar 4. 16 Visualisasi Label	36
Gambar 4. 17 Hasil Visualisasi Label	36
Gambar 4. 18 Splitting Data	37
Gambar 4. 19 Pembuatan Model	37
Gambar 4. 20 Smoothing	37
Gambar 4. 21 Confusion Matrix & Classification Report	38
Gambar 4. 22 Hasil Confusion Matrix & Classification Report	38
Gambar 4. 23 Visualisasi Plot ROC Curve	38
Gambar 4. 24 Hasil Visualisasi Plot ROC Curve	39
Gambar 4. 25 Pengujian Model	39
Gambar 4. 26 Hasil Pengujian Model Negatif	39
Gambar 4. 27 Hasil Pengujian Model Positif	40
Gambar 4. 28 Hasil Eksperimen	41

DAFTAR TABEL

Tabel 2. 1 Sumber Pustaka Primer	7
Tabel 2. 2 Confusion Matrix	15
Tabel 3. 1 Produk Online Shop	20
Tabel 4. 1. Pengumpulan Data	29
Tabel 4. 2. Labeling	29
Tabel 4. 3. Lower Teks	30
Tabel 4. 4. Normalisasi	30
Tabel 4. 5. Stopwords	31
Tabel 4. 6. Stemming	32
Tabel 4. 7. Hasil Uji Coba Parameter	40

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 *Shopee*. 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 resiko. Kondisi tersebut menuntut para pelaku *online shop* untuk dapat memahami kebutuhan dan keinginan yang semakin kompleks.

Analisis sentiment atau opinion mining merupakan analisa terhadap suatu pendapat seseorang pada sesuatu hal tertentu yang dituangkan kedalam sebuah pendapat, perkataan atau komentar untuk memberikan masukan tersendiri bagi pihak yang sedang di berikan opini. Pada proses penelitian yang akan dilakukan ini peneliti akan membagi pelabelan data kedalam 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).

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah bagaimana melakukan sentimen analisis pada ulasan customer terhadap suatu produk di *online shop* berbahasa Indonesia dengan menggunakan algoritma *Naïve Bayes Classifier* sehingga dapat membandingkan dan menentukan suatu produk di *online shop* yang memiliki kualitas yang terbaik dari hasil analisis sentimen positif dan negatif terhadap ulasan customer dengan metode tersebut.

1.3 Ruang Lingkup

Dalam penelitian ini terdapat beberapa ruang lingkup yang dibahas agar penyusunan dan pembahasan penelitian dapat dilakukan secara terarah sesuai dengan yang diharapkan. Antara lain sebagai berikut :

1. Data ulasan kepuasan customer terhadap suatu produk *online shop*.
2. Metode klasifikasi yang digunakan adalah *Naïve Bayes Classifier*.
3. Klasifikasi Sentimen positif dan negatif.

1.4 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah untuk memperoleh pengetahuan yang dapat memecahkan masalah yang sesuai dengan rumusan masalah diatas, tujuan yang hendak dicapai dalam penelitian ini yaitu untuk melakukan analisis sentimen terhadap data komentar kepuasan suatu produk dan juga mengetahui hubungan antara ulasan suatu produk dengan keputusan konsumen dalam melakukan pembelian di *online shop*.

Dengan tercapainya tujuan penelitian di atas, maka diharapkan akan bermanfaat bagi perusahaan, manfaat penelitian bagi perusahaan adalah dapat memberikan sumbangan informasi bagi perusahaan-perusahaan *e-commerce* di Indonesia tentang faktor-faktor yang mendorong keputusan pembelian konsumen, terutama dari faktor ulasan produk.

1.5 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 PRODUK APLIKASI

Bab ini membahas tentang prototipe sistem yang akan dibuat dan juga menjelaskan mengenai hasil, pembahasan hasil serta pengembangan ke Tugas Akhir dari penelitian Analisis sentiment review menggunakan algoritma Naïve Bayes Classifier.

BAB V 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

TINJAUAN PUSTAKA DAN TEORI

2.1 Tinjauan Pustaka

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 *fined 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 (th)	Metode	Hasil
1	Analisis Sentimen <i>Review Pelanggan E-COMMERCE</i> di Indonesia Menggunakan Algoritma <i>Naïve Bayes Classifier</i>	Tito Dwiki Armawan (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> BMKG Nasional	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	<u>Yang Diusulkan:</u> Analisis Sentiment Review Produc Online Shop Menggunakan Algoritma <i>Naïve Bayes Classifier</i> (2022)	Muhammad Aulia (2022)	<i>Naïve Bayes Classifier</i>	-

2.2 Teori

2.2.1 E-Commerce

Menurut Ahmadi dalam Marcel dan Astri (2018) *E-Commerce* merupakan aktifitas pembelian dan penjualan melalui jaringan internet dimana pembeli dan penjual tidak bertemu secara langsung, melainkan berkomunikasi melalui media internet. Menurut Shabur Mitfa dalam Novita Sari dkk (2017) *e-commerce* adalah suatu proses membeli dan menjual produk-produk secara elektronik oleh konsumen dan dari perusahaan ke perusahaan dengan komputer sebagai perantar transaksi bisnis. *E-commerce* adalah segala bentuk aktifitas transaksi jual beli barang ataupun jasa dengan menggunakan media elektronik.

Menurut Dita Novita (2022) *E-commerce* merupakan transaksi jual beli produk jasa dan informasi antar mitra bisnis melalui jaringan komputer yaitu internet. Internet merupakan “*a global network of computer network*” atau jaringan komputer yang sangat besar yang terbentuk dari jaringan-jaringan kecil yang ada diseluruh dunia yang saling berhubungan satu sama lain. Selain itu, *E-commerce* juga dapat diartikan sebagai suatu proses berbisnis dengan memakai teknologi elektronik yang menghubungkan antara perusahaan, konsumen dan masyarakat dalam bentuk transaksi elektronik dan pertukaran atau penjualan barang, servis, dan informasi secara elektronik.

E-commerce (electronic commerce) merupakan proses yang memungkinkan teknologi-teknologi berbasis situs internet yang memfasilitasi perniagaan/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 Apriadi & Saputra (2017), beberapa kelebihan dari *E-commerce* yang tidak dimiliki oleh transaksi bisnis yang dilakukan secara *offline*, hal tersebut antara lain :

1. Produk : Banyak jenis produk yang bisa dipasarkan dan dijual melalui internet seperti pakaian, kendaraan, barang elektronik, dll.
2. Tempat menjual produk : Tempat menjual yang dimaksud adalah melalui internet yang memerlukan domain dan hosting.

3. Cara menerima pesanan : *email*, telepon, sms, dll.
4. Cara pembayaran : *Credit card*, *debit card*, *paypal*, dan tunai.
5. Metode pengiriman: Pos Indonesia, JNE, Tiki, EMS.
6. *Customer Service: Email*, *contact us*, telepon, *Chatting*.

2.2.2 Analisis Sentiment

Menurut Wibowo & Jumiati (2018), analisis sentimen atau bisa disebut juga dengan *opinion minig* adalah bidang studi yang menganalisis sebuah pendapat, sentiment, sikap, penilaian, evaluasi, emosi orang terhadap suatu produk, organisasi, layanan, masalah, topik, individu, dan atributnya.

Menurut Aldi Bagus Sasmita (2022), analisis sentimen ialah salah satu jenis riset atau penelitian yang membahas tentang text mining. Terdapat perhitungan komputasional dari dua hal, yaitu opini dan perasaan yang terkandung dalam sebuah teks. Perhitungan komputasional yang dimaksud adalah analisis sentiment.

Pada analisis sentiment terdiri dari 3 jenis opini, yaitu opini positif, opini negative dan opini netral, sehingga dengan analysis sentimen perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui *feedback* masyarakat atau pun para ahli (Hafidz, 2021), (V.A. Safitri et al., 2019).

2.2.3 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 penelitian 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

2.2.4 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.

2.2.5 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 TF-IDF bobot akan ditemukan dalam persamaan berikut:

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

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

2.2.6 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*

2.2.7 *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.

2.2.8 Akurasi

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

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

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%.

2.2.9 Google Colab

Google Colab atau *Google Colaboratory* adalah sebuah executable document yang dapat digunakan untuk menyimpan, menulis, serta membagikan

program yang telah ditulis melalui *Google Drive* dengan code yang ada didalamnya menggunakan bahasa pemrograman *Python* dengan format “notebook”.

Google Colab bisa digunakan untuk beragam keperluan dan memiliki manfaat, diantaranya :

- *Built-in-library machine learning* yang lengkap
- Berbasis *cloud*, sehingga tidak memakan *space* dalam memori komputer
- data dalam *Google Colab* dapat diakses dan diedit dengan mudah
- mempermudah proses kolaborasi antar tim
- memiliki fitur **GPU** dan **TPU** yang dapat dimanfaatkan secara gratis.

2.2.10 *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 :

- *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.

- *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.

- *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.

2.2.11 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.

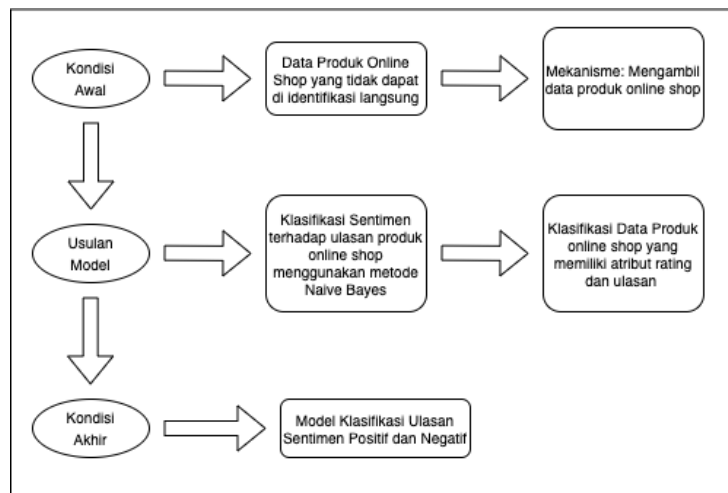
Visualisasi digunakan oleh penulis untuk mempermudah menjelaskan data yang akan dihasilkan. Agar para *e-commerce* dapat melihat dengan jelas dan tidak perlu banyak penjelasan untuk tau informasi atau data apa yang akan ditunjukkan.

BAB III

METODE PENELITIAN

3.1 Kerangka Penelitian

Pada tahap metodologi penelitian akan diuraikan mengenai metode yang digunakan dalam penelitian yang tercakup dalam kerangka penelitian. Kerangka penelitian dibuat untuk membuat pola pikir menjadi lebih realistis. Kerangka atau perancangan sistem dalam penelitian ini meliputi beberapa tahap, diantaranya pengumpulan data produk online shop, preprocessing, arsitektur model, pengklasifikasian data ulasan sentimen menggunakan algoritma *Naïve Bayes Classifier*. Kerangka penelitian disajikan pada gambar berikut.



Gambar 3. 1 Kerangka Penelitian

3.2 Data Penelitian

3.2.1 Sumber Data

Penelitian ini menggunakan data *product reviews dirty*. Data *e-commerce* yang digunakan berasal dari dataset public *Kaggle*.

Tabel 3. 1 Produk Online Shop

	text	rating	category	product_name	product_id	sold	shop_id	product_url
1	Barang sesuai pesanan dan cepat sampai	5	pertukangan	Staples Dekorasi Staples Kayu + Refill 8mm - S...	418660637	1	1740837	https://www.tokopedia.com/shakaonline87/staple...
2	Barang bagus harga murah	5	pertukangan	STAPLE GUN ATS 3 WAY TACKER - STAPLES JOK TEMB...	416032545	11	1477109	https://www.tokopedia.com/juraganperkakas/stap...
3	Paket rapi....mantap....cepat....sampe ke tujuan	5	pertukangan	STAPLE GUN ATS 3 WAY TACKER - STAPLES JOK TEMB...	416032545	11	1477109	https://www.tokopedia.com/juraganperkakas/stap...

3.2.2 Cara Mendapatkan Data

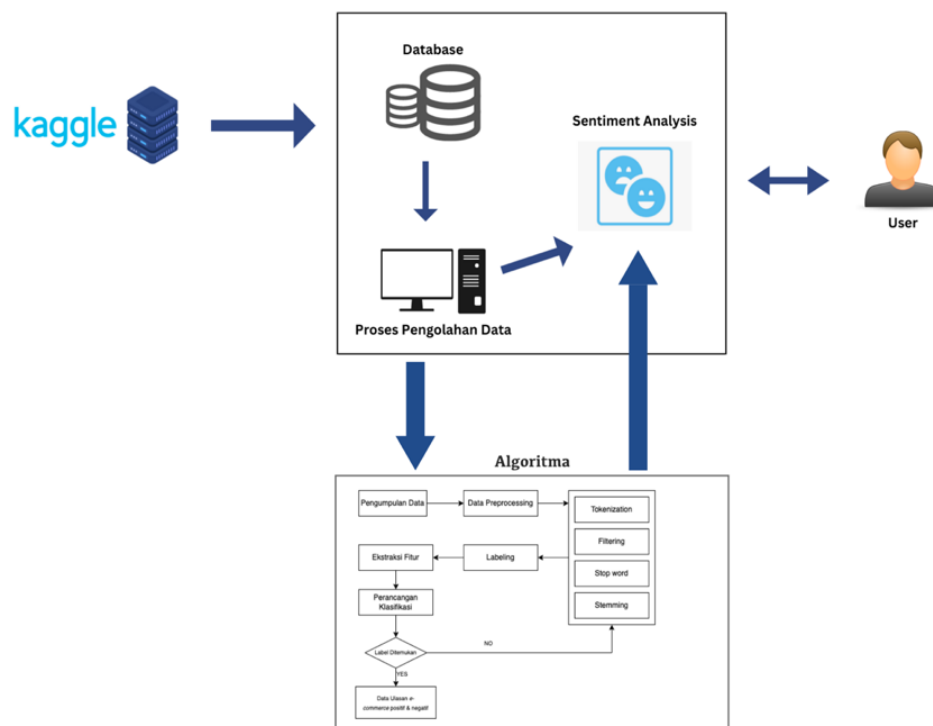
Dalam penelitian, data dikumpulkan dengan pengunduhan data pada situs dataset public *kaggle*. Data yang diunduh berukuran 10,5 MB pada dataset *product reviews dirty*. Dataset yang terkumpul memiliki 8 atribut.

3.2.3 Waktu Pengumpulan Data

Data diambil pada tanggal 25 September 2022.

3.3 Arsitektur Model

Sistem yang dibangun adalah sistem yang dapat digunakan untuk menganalisis sentimen pada ulasan produk dengan *sentiment analysis*, dikarenakan sebuah review atau ulasan dapat dinilai apabila dibaca secara keseluruhan. Sistem bekerja dengan melakukan pengambilan data dari sumber *Kaggle*, setelah itu sistem akan melakukan pengolahan pada data tersebut. Adapun pengguna dari sistem ini adalah user seperti gambar berikut.



Gambar 3. 2 Arsitektur Model

3.4 Analisis dan Perancangan

3.4.1 Kebutuhan Fungsional

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

2. Kebutuhan masukan

- **Data Produk Online Shop**

Data ini digunakan untuk mencatat persediaan barang dagang dan digunakan dalam mencatat transaksi yang berkaitan dengan barang dagang.

3. Kebutuhan proses

- **Sentiment Analysis Menggunakan Naïve Bayes**

Pada penelitian ini sistem melakukan proses menganalisis teks digital untuk menentukan apakah nada emosional pesan tersebut positif atau negatif.

4. Kebutuhan luaran

- **Model Prediksi**

Pada tahapan ini, hasil dari pengolahan data sudah dapat ditentukan apakah suatu ulasan produk *e-commerce* bersifat positif atau negatif.

3.4.2 Kebutuhan 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.

- **Google Colabotary**

Google Colabotary adalah *coding environment* bahasa pemrograman *python* dengan fromat “*notebook*”, atau dengan kata lain Google seakan meminjam komputer secara gratis untuk membuat program oleh Google. Google colab digunakan untuk *training dataset* dikarenakan spesifikasi komputer yang ditawarkan dengan spesifikasi tinggi (GPU Tesla, RAM-12, Disk-300GB), yang masih bisa sambung dengan Google Drive, sehingga mempercepat dalam proses *training dataset*.

2. Kebutuhan Perangkat Keras

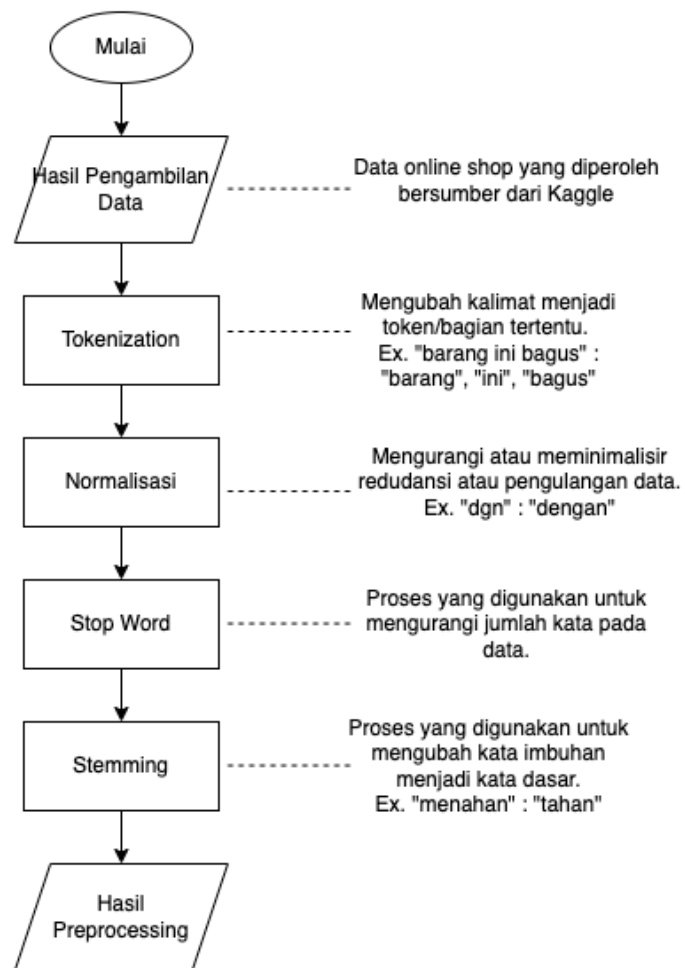
Perangkat keras yaitu peralatan komputer yang secara fisik yang berfungsi untuk melakukan analysis pada sebuah data. Sistem perangkat keras terdiri dari bagian input, bagian proses, bagian output dan bagian untuk melakukan percobaan aplikasi. Perangkat keras yang dibutuhkan ialah :

- 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.

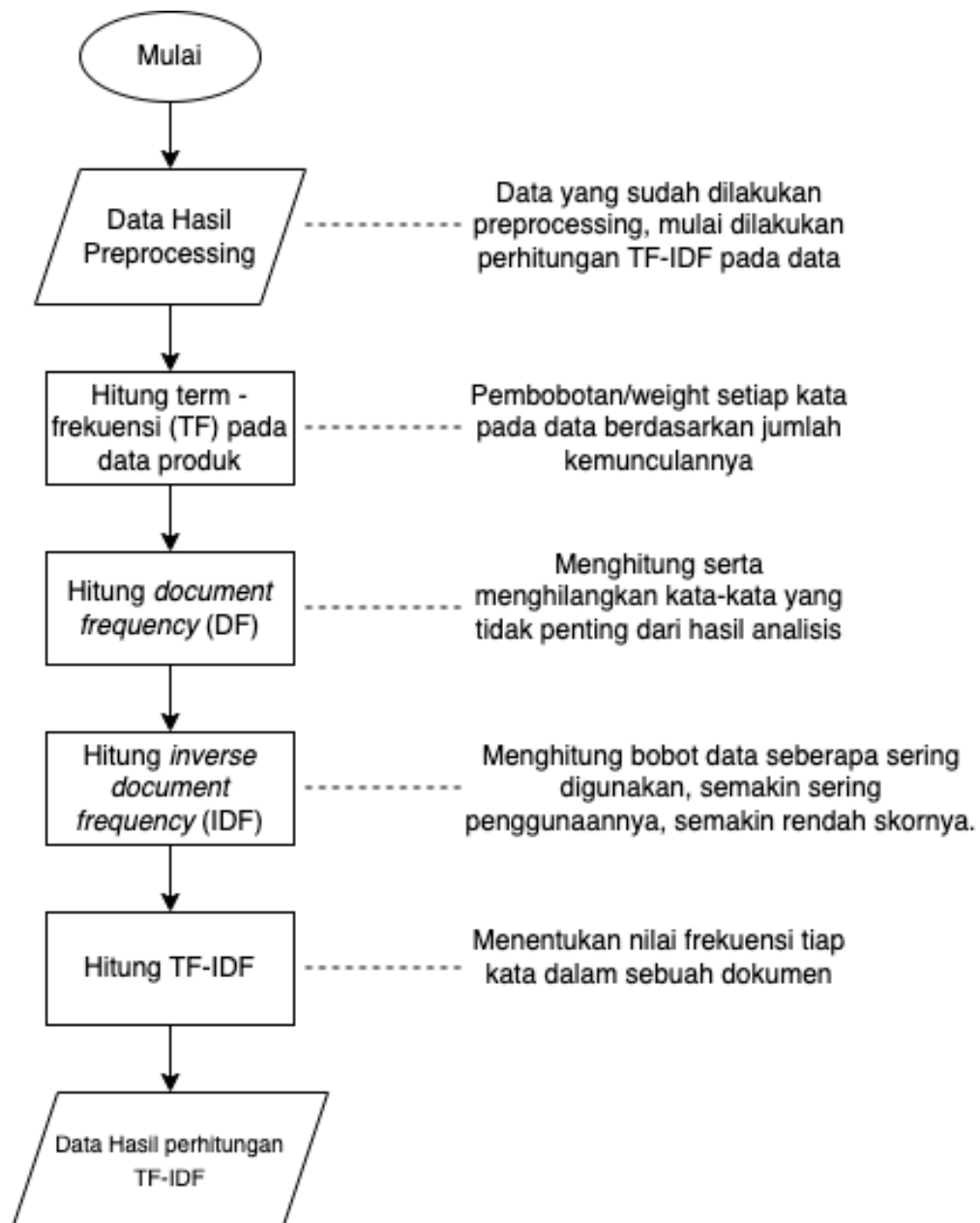
3.4.3 Perancangan Konseptual

Ada beberapa tahapan-tahapan yang akan dilalui dalam melakukan penelitian. Pertama yaitu pengambilan data, pada tahapan ini peneliti melakukan proses pengambilan data yang bersumber dari *Kaggle*. Selanjutnya, masuk ke tahapan prerocessing data yang dimana pada tahapan ini data yang sudah diperoleh akan dilakukan preprocessing sebelum masuk ke perhitungan TF-IDF. Komponen yang ada pada tahapan preprocessing diantaranya *tokenization*, *filtering*, *stop word* dan *stemming*. Tokenization digunakan untuk mengubah kalimat pada data menjadi token-token/bagian tertentu. Filtering digunakan untuk menyaring atau melakukan filter pada data tertentu. Stopword digunakan untuk mengurangi jumlah kata dalam sebuah dokumen. Stemming digunakan untuk memperkecil jumlah indeks yang

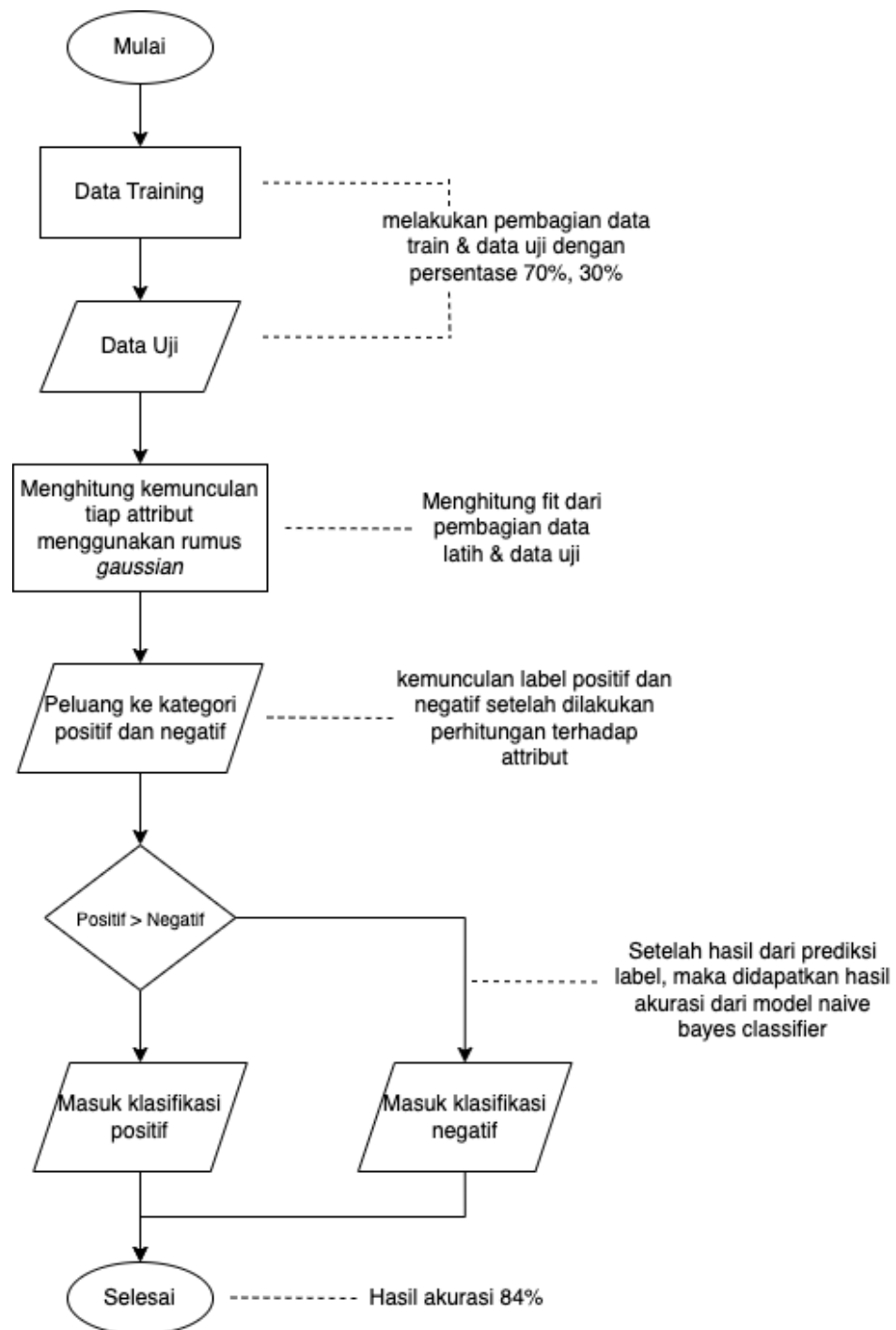
berbeda dari sebuah data. Tahapan kedua yaitu TF-IDF, pada tahapan ini peneliti akan melakukan serta menentukan nilai frekuensi sebuah data di dalam sebuah dokumen. Perhitungan TF-IDF ini juga menentukan seberapa relevan sebuah kata di dalam sebuah dokumen. Tahapan ketiga yaitu klasifikasi data, pada tahapan ini peneliti melakukan pembagian data menjadi data training dan data uji dengan random state 30, kemudian setelah dilakukan pembagian data peneliti menghitung kemunculan tiap-tiap atribut dengan menggunakan rumus *Gaussian* yang akan menghasilkan peluang untuk label positif dan negatif terhadap data, terakhir dari label yang diperoleh peneliti bisa melakukan klasifikasi dan pembagian kategori label untuk tiap-tiap data. Perancangan konseptual yang dilakukan dalam penelitian ini disajikan pada gambar berikut.



Gambar 3. 3 Preprocessing Data



Gambar 3. 4 TF-IDF



Gambar 3. 5 Klasifikasi Data

BAB IV

PRODUK

4.1 Hasil

Bagian ini menjabarkan proses ekstraksi yang digunakan untuk mengklasifikasi kata-kata yang bernilai positif dan negatif. Studi kasus yang diangkat adalah ulasan produk *e-commerce*. Akuisisi data teks berupa ulasan konsumen terhadap suatu produk yang ada pada toko online kemudian diekstraksi menjadi teks setelah itu diidentifikasi dan dikelompokkan menjadi kata-kata yang mengandung sentiment positif dan sentiment negatif. *Naïve Bayes Classifier* digunakan sebagai metode klasifikasi terhadap fitur ulasan pada attribut data. Pada penelitian ini data yang diambil yaitu data yang berupa teks kemudian diklasifikasikan berdasarkan kelas pada tiap kata tersebut. Adapun penjelasan detail dari tiap menu akan dijelaskan pada bagian berikutnya.

4.1.1 Tampilan Awal Prototipe

Adapun tahapan awal pada penelitian kali ini diantaranya :

- 1) Proses Pengumpulan Data
- 2) Labeling
- 3) Preprocessing Teks
- 4) TF-IDF
- 5) Visualisasi
- 6) Pembuatan Model
- 7) Pelatihan Model
- 8) Prototype Pengujian Model

4.1.2 Proses Pengumpulan Data

Tabel 4. 1. Pengumpulan Data

Unnamed: 0.1		Ulasan	Rating	Kategori	Nama Produk	Id Produk	Terjual	Id_Toko	Url
0	59	Pesanan gx sesuai dgn yg d gambar	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...
1	70	Pengiriman dan respon super cepat. Tpi barang ...	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...
2	95	barang yg dikirim tidak sesuai pesanan	1	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...
3	117	Php, bilang isi ada disekripsi pas dipesen be...	1	pertukangan	steples tembak / staples tembak kenmaster PROM...	88842566	45	1102298	https://www.tokopedia.com/cahayabelawa/steples...
4	153	barang jelek copot" terus tmpat taro isinya	2	pertukangan	Staple Gun / Staples Tembak / Staples Jok / He...	209226141	171	580197	https://www.tokopedia.com/hmhhardware/staple-g...

Pada Tabel 4.1 dijelaskan dokumen data yang digunakan yaitu dokumen berlabel teks ulasan *e-commerce* kemudian dokumen tersebut dimasukkan sebagai proses pembelajaran. Pada tahapan ini data yang digunakan masih utuh atau belum dibersihkan dan belum diberikan label pada tiap-tiap ulasan sehingga tiap kata di teks tersebut masih bercampur dengan karakter-karakter lainnya yang masih melekat pada data tersebut. data teks yang akan diolah masih bercampur.

4.1.3 Labeling

Tabel 4. 2. Labeling

Ulasan	Rating	Kategori	Nama Produk	Id Produk	Terjual	Id_Toko	Url	label
Pesanan gx sesuai dgn yg d gambar	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
Pengiriman dan respon super cepat. Tpi barang ...	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
barang yg dikirim tidak sesuai pesanan	1	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
Php, bilang isi ada disekripsi pas dipesen be...	1	pertukangan	steples tembak / staples tembak kenmaster PROM...	88842566	45	1102298	https://www.tokopedia.com/cahayabelawa/steples...	0
barang jelek copot" terus tmpat taro isinya	2	pertukangan	Staple Gun / Staples Tembak / Staples Jok / He...	209226141	171	580197	https://www.tokopedia.com/hmhhardware/staple-g...	0

Pada Tabel 4.2 peneliti melakukan pelabelan untuk tiap-tiap baris data berdasarkan attribut *Rating* yang ada pada data. Nilai label yang diberikan yaitu 0 dan 1 yang dimana label dari 0 merupakan sentiment negatif, sedangkan label 1 merupakan sentiment positif. Script dari labeling ditunjukkan pada gambar di bawah.

```
[ ] label = []
for index, row in df_copy.iterrows():
    if row["Rating"] == 1 or row["Rating"] == 2:
        label.append(0)
    else :
        label.append(1)

df_copy["label"] = label
```

Gambar 4. 1 Labeling

4.1.4 Preprocessing Teks

Tahapan pengolahan data asli yang telah diolah berupa data teks sebelum data tersebut diolah lebih lanjut. Tujuan dari preprocessing teks adalah untuk menghilangkan noise, memperjelas fitur, mengkonversi data asli agar sesuai kebutuhan serta memperbesar atau memperkecil data. Dalam preprocessing memiliki beberapa tahapan, diantaranya :

- **Lower Teks**

Tabel 4. 3. Lower Teks

Ulasan	Rating	Kategori	Nama Produk	Id Produk	Terjual	Id_Toko	Url	label
pesanan gx sesuai dgn yg d gambar	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
pengiriman dan respon super cepat. tpi barang ...	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
barang yg dikirim tidak sesuai pesanan	1	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
php, bilang isi ada disekripsi pas dipesen be...	1	pertukangan	steples tembak / staples tembak kenmaster PROM...	88842566	45	1102298	https://www.tokopedia.com/cahayabelawa/steples...	0
barang jelek copot" terus tmpat taro isinya	2	pertukangan	Staple Gun / Staples Tembak / Staples Jok / He...	209226141	171	580197	https://www.tokopedia.com/hmhhardware/staple-g...	0

Proses ini merupakan tahapan yang dimana keseluruhan attribute ulasan yang ada pada dokumen diubah menjadi huruf kecil (lower). Script dari lower ditunjukkan pada gambar di bawah.

```
[ ] # Mengubah text menjadi lower
df_copy['Ulasan'] = df_copy['Ulasan'].str.lower()
```

Gambar 4. 2 Lower Teks

- **Normalisasi**

Tabel 4. 4. Normalisasi

Ulasan	Rating	Kategori	Nama Produk	Id Produk	Terjual	Id_Toko	Url	label
pesanan tidak sesuai dengan yang di gambar	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
pengiriman dan respon super cepat. tapi barang...	2	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
barang yang dikirim tidak sesuai pesanan	1	pertukangan	Isi Staples Tembak 8 mm best guard	133507638	545	1461393	https://www.tokopedia.com/timurjaya46/isi-stap...	0
php, bilang isi ada disekripsi saat dipesen b...	1	pertukangan	steples tembak / staples tembak kenmaster PROM...	88842566	45	1102298	https://www.tokopedia.com/cahayabelawa/steples...	0
barang jelek copot" terus tmpat taro isinya	2	pertukangan	Staple Gun / Staples Tembak / Staples Jok / He...	209226141	171	580197	https://www.tokopedia.com/hmhhardware/staple-g...	0

Proses normalisasi merupakan proses yang digunakan untuk melakukan normalisasi serta mengubah format teks untuk memenuhi tujuan tertentu. Script dari normalisasi ditunjukkan pada gambar di bawah.

```
[ ] norm = {"dgn": "dengan", "gue": "saya", "bgmn": "bagaimana", "tdk": "tidak", "blum": "belum", "mantaaaaaapp": "bagus", "joss": "joss", "blum": "belum", "mantaaaaaapp": "bagus", "joss": "joss", "thanks": "terima kasih", "fast": "cepat", "trim": "terima kasih", "brg": "barang", "gx": "tidak", "dgn": "dengan", "recommended": "rekomendasi", "recomend": "rekomendasi", "good": "bagus", "eksis": "ada", "beenilai": "bernilai", "dg": "dengan", "yg": "yang", "t4": "tempat", "kl": "kalo", "k": "ke", "plg": "pulang", "ajah": "aja", "bgt": "banget", "lbh": "lebih", "ayem": "tenang", "dsana": "disana", "lg": "lagi", "pas": "saat", "nggak": "tidak", "karna": "karena", "utk": "untuk", "dn": "dan", "mlht": "melihat", "pd": "pada", "mndgr": "mendengar", "crita": "cerita", "sejrh": "sejarah", "mmbh": "menambah", "sayapun": "saya", "thn": "tahun", "halu": "halusinasi", "nyantai": "santai", "plus": "dan", "ayang": "saya", "Rekomended": "direkomendasikan", "now": "sekarang", "slalu": "selalu", "photo": "foto", "slah": "salah", "krn": "karena", "aya": "saya", "slip": "bagus", "bny": "banyak", "tdk": "tidak", "byk": "banyak", "pool": "sekali", "pgn": "ingin", "gue": "saya", "bgmn": "bagaimana", "ga": "tidak", "gak": "tidak", "dr": "dari", "yg": "yang", "lu": "kamu", "sya": "saya", "lancarr": "lancar", "kayak": "seperti", "ngawur": "sembarangan", "k": "ke", "luass": "luas", "sy": "saya", "thn": "tahun", "males": "malas", "tgl": "tanggal", "lg": "lagi", "bgt": "banget", "gua": "saya", "n": "\n", "tpi": "tapi", "standar": "biasa", "standart": "biasa", "sdh": "sudah"}

def normalisasi(str_text):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text

df_copy['Ulasan'] = df_copy['Ulasan'].apply(lambda x: normalisasi(x))
```

Gambar 4. 3 Normalisasi

- **Stopwords**

Tabel 4. 5. Stopwords

Ulasan	Rating	Kategori	Nama Produk	Id Produk	Terjual	Id_Toko	Url	label
bagus memuaskan. packing bagus cepat	4	pertukangan	Staples Gun Tacker Mollar 3 in 1 / Staple Jok ...	221298191	408	1114588	https://www.tokopedia.com/indahjayatools/stapl...	1
mantap pengiriman oke cepat rekomendasi	5	pertukangan	Staples Gun Tacker Mollar 3 in 1 / Staple Jok ...	221298191	408	1114588	https://www.tokopedia.com/indahjayatools/stapl...	1
pengiriman cepat. lumayan	4	pertukangan	Staples Gun Tacker Mollar 3 in 1 / Staple Jok ...	221298191	408	1114588	https://www.tokopedia.com/indahjayatools/stapl...	1
belum dicoba semoga oke barangnya	4	pertukangan	Staples Gun Tacker Mollar 3 in 1 / Staple Jok ...	221298191	408	1114588	https://www.tokopedia.com/indahjayatools/stapl...	1
barang cukup bagus, respon bagus. manualnya ga...	3	pertukangan	Staples Gun Tacker Mollar 3 in 1 / Staple Jok ...	221298191	408	1114588	https://www.tokopedia.com/indahjayatools/stapl...	1

Pada tabel 4.5 dilakukan tahapan stopwords pada teks dengan tujuan untuk meminimalisir jumlah kemunculan kata yang dianggap tidak memiliki makna. Script dari stopwords ditunjukkan pada gambar di bawah.

```
[ ] from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary
more_stop_words = []

stop_words = StopWordRemoverFactory().get_stop_words()
new_array = ArrayDictionary(stop_words)
stop_words_remover_new = StopWordRemover(new_array)

def stopword(str_text):
    str_text = stop_words_remover_new.remove(str_text)
    return str_text

df_copy['Ulasan'] = df_copy['Ulasan'].apply(lambda x: stopword(x))
df_copy.tail()
```

Gambar 4. 4 Stopwords

- **Tokenize**

Pada tahapan ini dilakukan tokenize pada teks yang ada di atribut ulasan, dengan tujuan untuk memisahkan tiap kata yang ada pada kalimat. Script dan hasil dari tokenize ditunjukkan pada gambar di bawah.

```
[ ] tokenized = df_copy['Ulasan'].apply(lambda x:x.split())
tokenized
```

Gambar 4. 5 Tokenize

```
0          [pesanan, sesuai, yang, gambar]
1  [pengiriman, respon, super, cepat., barang, se...
2          [barang, dikirim, sesuai, pesanan]
3  [php,, bilang, isi, disekripsi, dipesen, berb...
4  [barang, jelek, copot&#34;, terus, tmpat, taro...
...
995          [bagus, memuaskan., packing, bagus, cepat]
996          [mantap, pengiriman, oke, cepat, rekomen, ed]
997          [pengiriman, cepat., lumayan]
998          [belum, dicoba, semoga, oke, barangnya]
999  [barang, cukup, bagus,, respon, bagus., manual...
Name: Ulasan, Length: 1925, dtype: object
```

Gambar 4. 6 Hasil Tokenize

- **Stemming**

Tabel 4. 6. Stemming

index	Ulasan	label
0	pesan sesuai yg d gambar	0
1	kirim respon super cepat tpi barang sesuai gambar tidak kuat tembus jok motor	0
2	barang yg kirim sesuai pesan	0
3	php bilang isi skripsi pas sen beda mubajir rp50 000 nyata isi ga cocok	0
4	barang jelek copot 34 terus tmpat taro isi	0
5	beli staples gak jual isi sekali pake ckckck	0
6	baik kalau mau ngirim barang periksa dulu gan kondisi barang layak apa gak gitu loh isi karat semua	0
7	produk sesuai order prohex yg kirim	0
8	barang sdh terima	0
9	bisa buat masang jok gak pake	0

Pada tabel 4.6 dijelaskan tahapan mengubah kata menjadi kata dasar dengan menghilangkan imbuhan (*affixes*) yaitu awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan confixes (kombinasi dari awalan dan akhiran) pada kata turunan yang ada di dokumen akan dicocokkan dengan KBBI dan diubah sesuai pada aturan pada stemming dan apabila dicek di kamus diubah sesuai aturan kata tidak ditemukan sebagai kata dasar, maka kata tersebut dikembalikan ke bentuknya semula dan

dihitung sebagai kata dasar baru. Script dari stemming ditunjukkan pada gambar di bawah.

```
[19] from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(Ulasan):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in Ulasan:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

tokenized.to_csv('databersih2.csv', index=False)
data_clean = pd.read_csv('databersih2.csv', encoding='latin1')
data_clean.head()
```

Gambar 4. 7 Stemming

4.1.5 TF-IDF

Pada tahapan ini dijelaskan untuk menghitung dan informasi TF (*term frequency*), DF (*document frequency*) dan IDF (*inverse document frequency*) menghitung dokumen atau term ini berdasarkan frekuensi kemunculan term atau dokumen tersebut. Kata/term dihitung probabilitas kemunculan dalam satu dokumen. Untuk mendapatkan IDF digunakan persamaan dengan menggunakan persamaan. Script dan hasil dari TF-IDF ditunjukkan pada gambar di bawah.

```
[ ] from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
Ulasan = result['Ulasan']
cv = CountVectorizer()
term_fit = cv.fit(Ulasan)
term_fit.vocabulary_
```

Gambar 4. 8 Menghitung Kata

```
'gojek': 961,
'direfund': 758,
'info': 1085,
'belum': 362,
'dinyalain': 752,
'alarm': 175,
...
```

Gambar 4. 9 Hasil Perhitungan Kata

```
[28] # Kolom pertama ini berarti jumlah dokumen
      # Kolom kedua berarti letak kata nya
      # Kolom ketiga berarti hasil dari tf

term_frequency_all = term_fit.transform(Ulasan)
print(term_frequency_all)
```

Gambar 4. 10 Proses TF

```
(0, 913)      1
(0, 1979)     1
(0, 2319)     1
(0, 2840)     1
(1, 301)      1
```

Gambar 4. 11 Hasil TF

4.1.6 Visualisasi

Pada tahapan ini data yang sudah dilakukan pelabelan dan preprocessing akan ditampilkan menggunakan visualisasi untuk tiap-tiap kata yang mengandung sentiment positif dan sentiment negatif. Visualisasi ini menampilkan tiap kata yang sering muncul pada teks serta visualisasi jumlah kata yang memiliki label positif dan negatif. Script dan hasil dari tahapan visualisasi ditunjukkan pada gambar di bawah.

```
[36] all_text_s0 = ' '.join(word for word in train_s0["Ulasan"])
wordcloud = WordCloud(colormap='Reds', width=1000, height=1000, mode='RGBA', background_color='white').generate(all_text_s0)
plt.figure(figsize=(20, 10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Ulasan Negatif")
plt.margins(x=0, y=0)
plt.show()
```

Gambar 4. 12 Visualisasi Negatif



Gambar 4. 13 Hasil Visualisasi Negatif

Hasil dari visualisasi negatif pada gambar di atas menampilkan kemunculan kata yang memiliki label negatif (0) berdasarkan dari analisis yang dilakukan.

```
[40] all_text_s1 = ' '.join(word for word in train_s1["Ulasan"])
wordcloud = WordCloud(colormap='Blues', width=1000, height=1000, mode="RGBA", background_color='white').generate(all_text_s1)
plt.figure(figsize=(20, 10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Ulasan Positif")
plt.margins(x=0, y=0)
plt.show()
```

Gambar 4. 14 Visualisasi Positif

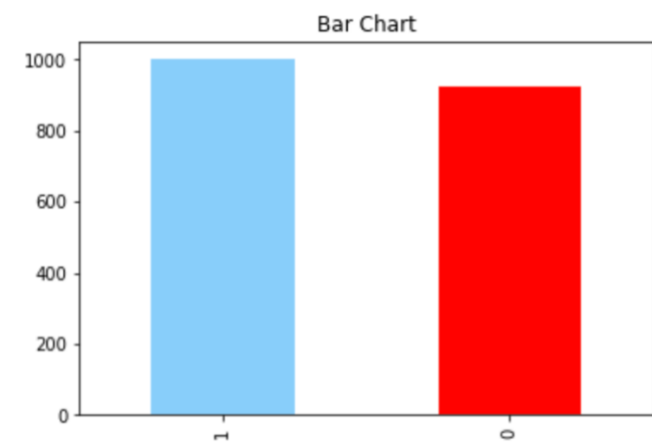


Gambar 4. 15 Hasil Visualisasi Positif

Hasil dari visualisasi positif pada gambar di atas menampilkan kemunculan kata yang memiliki label positif (1) berdasarkan dari analisis yang dilakukan.

```
[41] sentimen_data = pd.value_counts(df_copy["label"], sort=True)
sentimen_data.plot(kind='bar', color=['lightskyblue', 'red'])
plt.title("Bar Chart")
plt.show
```

Gambar 4. 16 Visualisasi Label



Gambar 4. 17 Hasil Visualisasi Label

Berdasarkan hasil visualisasi label pada gambar di atas, label positif menghasilkan chart yang sedikit mendominasi dibandingkan dengan label negatif,

artinya kemunculan kata yang memiliki label positif lebih banyak dibandingkan label negatif.

4.1.7 Pembuatan Model

Pada tahapan ini akan dilakukan split data terlebih dahulu secara *random* dengan data latih 70% dan data uji 30% yang akan diklasifikasikan menggunakan model pembelajaran *Naïve Bayes Classifier*. Script ditunjukkan pada gambar di bawah.

```
[43] from sklearn.model_selection import train_test_split

      X_train, X_test, y_train, y_test = train_test_split(result['Ulasan'], result['label'],
                                                         test_size=0.5, stratify=result['label'], random_state=30)
```

Gambar 4. 18 Splitting Data

Setelah dilakukan splitting data pada gambar di atas, kemudian peneliti melakukan inisialisasi pembuatan model *Naïve Bayes Classifier*. Script ditunjukkan pada gambar di bawah.

```
[51] from sklearn.naive_bayes import GaussianNB

      nb = GaussianNB()
      model = nb.fit(vec_data, y_train)
```

Gambar 4. 19 Pembuatan Model

4.1.8 Pelatihan Model (*Naïve Bayes Classifier*)

Pada tahapan ini peneliti melakukan pelatihan pada model yang sudah di inisialisasikan dengan melakukan smoothing pada model terlebih dahulu dan mendapatkan hasil dari proses smoothing tersebut sebelum dilakukan perhitungan selanjutnya terhadap model. Script dan hasil ditunjukkan pada gambar di bawah.

```
[55] nb.fit(X_train, y_train)

      GaussianNB(var_smoothing=0.0054622772176843425)
```

Gambar 4. 20 Smoothing

Sesudah melakukan proses smoothing terhadap X_{train} dan y_{train} , kemudian masuk ke tahapan perhitungan *Precision* dan *confusion matrix*. Script dan hasil yang didapatkan berdasarkan perhitungan tersebut ditunjukkan pada gambar di bawah.

```
[57] from sklearn.metrics import confusion_matrix
      from sklearn.metrics import classification_report, roc_curve, plot_roc_curve

[58] print('----- confusion matrix -----')
      print(confusion_matrix(y_test, y_pred_nb))

      print('----- classification report -----')
      print(classification_report(y_test, y_pred_nb))
```

Gambar 4. 21 Confusion Matrix & Classification Report

```
----- confusion matrix -----
[[370  93]
 [ 61 439]]
----- classification report -----
```

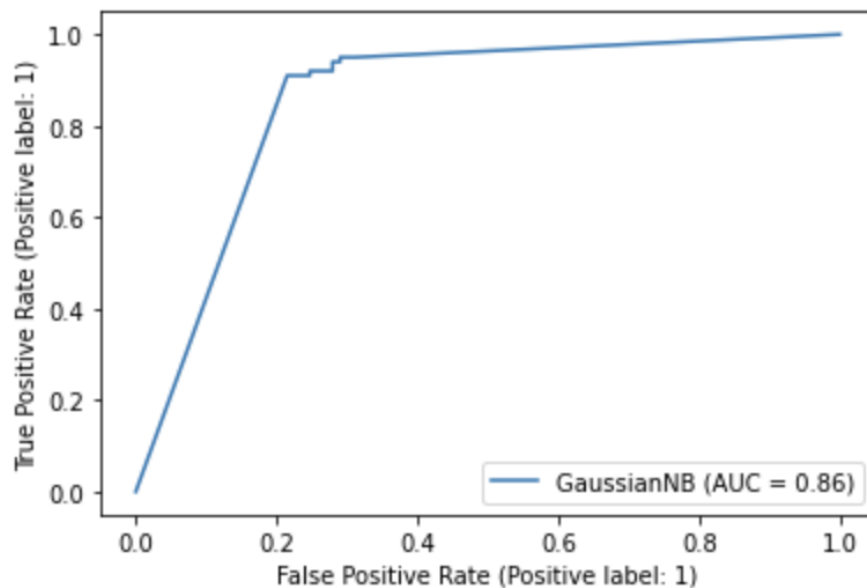
	precision	recall	f1-score	support
0	0.86	0.80	0.83	463
1	0.83	0.88	0.85	500
accuracy			0.84	963
macro avg	0.84	0.84	0.84	963
weighted avg	0.84	0.84	0.84	963

Gambar 4. 22 Hasil Confusion Matrix & Classification Report

Berikutnya, setelah mendapatkan hasil dari perhitungan *confusion matrix* & *classification report* dengan nilai akurasi yaitu 84%, nilai tersebut akan dilakukan visualisasi *plot_ruc_curve*. Script dan hasil ditunjukkan pada gambar di bawah.

```
[59] plot_roc_curve(nb, X_test, y_test)
```

Gambar 4. 23 Visualisasi Plot ROC Curve



Gambar 4. 24 Hasil Visualisasi Plot ROC Curve

4.1.9 Prototype Pengujian Model

Nilai akurasi yang didapatkan pada hasil perhitungan yaitu 84%, kemudian peneliti melakukan pengujian menggunakan input data baru dengan tujuan untuk mengetahui sentiment dari input data tersebut. Script dan hasil dari pengujian ditunjukkan pada gambar yang ada di bawah.

```
[60] df_pred = input(str("Input Text : "))
      teks = [df_pred]

      df_pred = pd.DataFrame(teks, columns=["text"])
      df_pred
      #Prediksi/Klasifikasi
      df_pred["Sentiment"] = model.predict((vec_model.transform(df_pred['text']).toarray()))
      df_pred["Sentiment"] = df_pred["Sentiment"].replace(to_replace=[0, 1], value=["NEGATIF", "POSITIF"])
      df_pred
```

Gambar 4. 25 Pengujian Model

Input Text : di foto bagus, pas barang sampai jelek

text

Sentiment

0

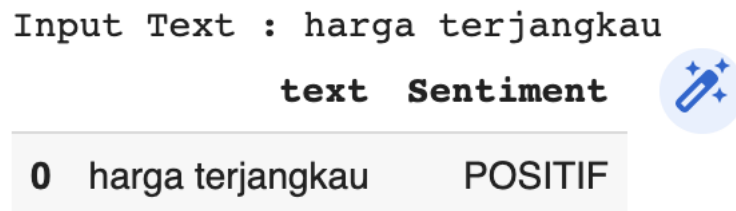
di foto bagus, pas barang sampai jelek

NEGATIF

Gambar 4. 26 Hasil Pengujian Model Negatif

Gambar 4.26. Hasil Pengujian Model Negatif

Berdasarkan gambar di atas, hasil dari pengujian model yang diinputkan memberikan label sentiment *negatif*.



Gambar 4. 27 Hasil Pengujian Model Positif

Berdasarkan gambar di atas, hasil dari pengujian model yang diinputkan memberikan label sentiment *positif*.

4.2 Pembahasan Hasil

Prototype yang dihasilkan digunakan untuk melakukan klasifikasi ulasan sentiment pada *e-commerce*. Adapun eksperimen dengan menggunakan beberapa parameter dari klasifikasi ulasan sentiment terhadap suatu produk di *e-commerce* seperti pada tabel di bawah.

Tabel 4. 7. Hasil Uji Coba Parameter

No	n_splits	n_repeats	random_state	logspace	verbose	var_smoothing	precision		Recall		F1-Score		Accuracy
							0	1	0	1	0	1	
1	5	3	999	0, -9, 100	1	0.0023101297000831605	84%	81%	79%	86%	81%	84%	82%
2	7	5	599	0, -7, 200	1	0.0031806256927941192	85%	82%	79%	87%	82%	84%	83%
3	3	6	700	0, -7, 100	1	0.0054622772176843425	86%	83%	80%	88%	83%	85%	84%
4	8	5	999	0, -7, 250	1	0.0016476595119256485	82%	81%	78%	85%	80%	83%	82%

Seperti pada tabel diatas, hasil tes dari uji coba pada berbagai perubahan parameter menampilkan hasil *akurasi* tertinggi yaitu 84%. Hal ini dikarenakan proses smoothing yang dilakukan memberikan hasil *var_smoothing* tertinggi dibandingkan dengan uji coba parameter lainnya. Kemudian hasil dari classification report yang diperoleh juga memberikan akurasi yang cukup tinggi dibandingkan yang lainnya. Adapun hasil dari uji coba parameter dengan akurasi tertinggi seperti pada gambar di bawah.

```

----- confusion matrix -----
[[370  93]
 [ 61 439]]
----- classification report -----

```

	precision	recall	f1-score	support
0	0.86	0.80	0.83	463
1	0.83	0.88	0.85	500
accuracy			0.84	963
macro avg	0.84	0.84	0.84	963
weighted avg	0.84	0.84	0.84	963

Gambar 4. 28 Hasil Eksperimen

4.3 Pengembangan ke Tugas Akhir

Berdasarkan pembahasan hasil prototype yang sudah dilakukan, diketahui bahwa prototype masih memiliki beberapa kekurangan, diantaranya ada pada bagian normalisasi teks, karena tiap-tiap ulasan pasti mengandung beberapa kalimat yang ambigu, oleh sebab itu diperlukan normalisasi yang lebih detail pada tiap kata. Selanjutnya pada visualisasi, dikarenakan tahapan visualisasi sangat bergantung pada hasil dari preprocessing, maka dari itu perlu diperhatikan untuk kata yang ditampilkan pada visualisasi sentiment positif dan sentiment negatif, pada prototype visualisasi masih ada beberapa kata yang muncul pada sentiment yang berlawanan, oleh sebab itu pengembangan pada tahapan preprocessing perlu dilakukan. Terakhir pada testing, masih ada beberapa kata yang memiliki sentiment yang berlawanan, akibatnya pada pengujian teks, label sentiment yang dihasilkan masih belum 100% akurat. Dari beberapa kekurangan yang sudah dijelaskan di atas, peneliti menentukan arah pengembangan di Tugas Akhir sebagai berikut :

- 1) Melakukan pengembangan terhadap preprocessing teks.
- 2) Melakukan riset ulang terhadap data ulasan produk yang di *e-commerce*.
- 3) Membuat interface terhadap tahap pengujian model.
- 4) Menambahkan hasil output yaitu atribut rating pada suatu produk.

BAB V

SIMPULAN

Berdasarkan keseluruhan proses analisis, perancangan, dan implementasi pada pembuatan sistem analisis sentimen produk *e-commerce* menggunakan metode *naïve bayes*, diperoleh kesimpulan sebagai berikut :

1. Sistem dapat mendeteksi sentiment positif dan negatif dari ulasan produk *e-commerce* dengan menerapkan metode *Naïve Bayes Classifier*.
2. Pada penelitian ini, algoritma *Naïve Bayes Classifier* terbukti merupakan algoritma yang akurat karena tingkat keakuratan sistem analisis sentimen produk *e-commerce* menggunakan algoritma *naïve bayes* yang diuji menggunakan data uji mendapatkan hasil akurasi sebesar **84%** dengan tingkat kebenaran 1000 untuk label positif dan 925 untuk label negatif.
3. Kelebihan pada project ini adalah memiliki nilai akurasi, presisi dan *recall* yang baik, sehingga sudah cukup untuk bisa digunakan dalam sebuah sistem.
4. Kekurangan pada project ini terdapat pada kinerja sistem dalam mencari keberhasilan untuk menemukan sebuah informasi tiap-tiap kata pada kelas *positif* dan *negatif*. Hal itu disebabkan karena proses pembobotan vectorizer pada tiap kata dengan label tertentu memiliki nilai bobot yang cenderung lebih tinggi, akibatnya ada beberapa kalimat akan memiliki label sebaliknya dari label yang seharusnya.

DAFTAR PUSTAKA

- Alzami, F., P. N. P., Anggi, R., Megantara, R. A., & Prabowo, D. P. (2020). Sentiment Analysis Untuk Deteksi Ujaran Kebencian Pada Domain Politik. *Science And Engineering National Seminar*, 213-218.
- Armawan, T. D. (2022). Analisis Sentimen Review Pelanggan E-commerce di Indonesia Menggunakan Algoritma Naive Bayes Classifier.
- Chaudhuri, K. D. (2022, March 21). Building Naive Bayes Classifier from Scratch to Perform Sentiment Analysis. Retrieved from Analytics Vidhya: <https://www.analyticsvidhya.com/blog/2022/03/building-naive-bayes-classifier-from-scratch-to-perform-sentiment-analysis/>
- Darwis, D., Siskawati, N., & Abidin, Z. (2020). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal Tekno Kompak*, 131-145.
- Janah, A. K., Wahyuni, E. D., & Arifiyanti, A. A. (2020). Klasifikasi Emosi Ulasan Aplikasi Traveloka Pada Google Play Menggunakan Naive Bayes. *Jurnal Informatika dan Sistem Informasi*, 1716-1722.
- Mesri, A. (2022, March 23). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. Retrieved from Medium: <https://blog.devgenius.io/twitter-sentiment-analysis-hate-speech-detection-with-unbalanced-data-46d1f57be2ab>
- Qamal, M., Fuadi, W., & Muslia. (n.d.). Analisis Sentimen Toko Online Menggunakan Algoritma Naive Bayes Classifier. 124-132.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *JURNAL INOVTEK POLBENG*.
- Yulita, W., Nugroho, E. D., & Algifari, M. H. (2020). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier. *JDMSI*.