

**IMPLEMENTASI ALGORITMA KLASIFIKASI *NAIVE BAYES*
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP CHATGPT**

SKRIPSI

**M RIEKI NURUL RA
20190050042**



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN**

SUKABUMI

JULI 2023

**IMPLEMENTASI ALGORITMA KLASIFIKASI *NAIVE BAYES*
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP CHATGPT**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam
Menempuh Gelar Sarjana Komputer*

M. RIEKI NURUL RA
20190050042



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN**

SUKABUMI

JULI 2023



PERNYATAAN PENULIS

JUDUL : IMPLEMENTASI ALGORITMA KLASIFIKASI *NAÏVE BAYES*
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP CHATGPT
NAMA : M RIFKI NURUL R A
NIM : 20190050042

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti - bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut.”

Sukabumi, 3 Juli 2023



M Rifki Nurul R A
20190050042

PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI ALGORITMA KLASIFIKASI NAÏVE BAYES
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP CHATGPT
NAMA : M RIFKI NURUL R A
NIM : 20190050042

“Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal 3 Juli 2023 Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom)”

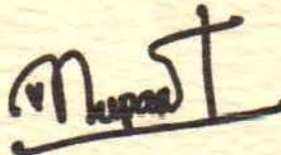
Sukabumi, 3 Juli 20203

Pembimbing I



Adhitia Erfina, S.T.,M.Kom
NIDN. 0417049102

Pembimbing II



Mupaat, M.Kom
NIDN. 0414027906

Ketua Penguji



Dede Sukmawan, M.Kom
NIDN. 0405119501

Ketua Program Studi Sistem informasi



Adhitia Erfina, S.T.,M.Kom
NIDN. 0417049102

Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun, S.T., M.T., IPM., Asean Eng.
NIDN. 0402037401

ABSTRACT

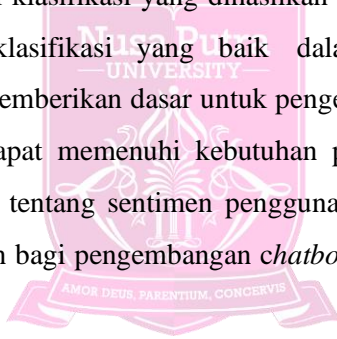
ChatGPT (Generative Pre-Trained Transformer) is a chatbot that is being widely used by the public. This technology is based on Artificial Intelligence and is capable of having conversational interactions with its users just like humans, but in the form of automated text. Because of this capability, online forums such as Brainly and the like can be overtaken by these smart chatbots. Therefore, this study was conducted to determine the positive and negative sentiments towards ChatGPT using Naive Bayes Classification algorithm on 5000 Twitter users. Data was collected by scraping technique and Python programming language was used in data analysis. The results showed that the majority of Twitter users had a positive sentiment of 57.6% towards ChatGPT, while the negative sentiment reached 42.4%. The resulting classification model had an accuracy of 80%, indicating a good classification model in determining sentiment probabilities. These findings provide a basis for the development of better AI chatbot technology that can meet user needs. The results of this study provide insights into user sentiment towards ChatGPT and can be used as a reference for future AI chatbot development.

Keywords: *ChatGPT, Sentiment Analysis, Naive Bayes Classifier.*



ABSTRAK

ChatGPT (*Generative Pre-Trained Transformer*) adalah chatbot yang sedang marak digunakan oleh masyarakat. Teknologi ini berbasis *Artificial Intelligence* (kecerdasan buatan) dan mampu melakukan interaksi percakapan dengan penggunanya seperti layaknya manusia, namun dalam bentuk teks otomatis. Karena kemampuan ini, forum-forum online seperti Brainly dan sejenisnya dapat tersisihkan oleh chatbot pintar ini. Oleh karena itu, penelitian ini dilakukan untuk mengetahui sentimen positif dan negatif terhadap ChatGPT dengan menggunakan algoritma Klasifikasi Naive Bayes pada 5000 pengguna Twitter. Data dikumpulkan dengan teknik *scraping* dan bahasa pemrograman Python digunakan dalam analisis data. Hasil penelitian menunjukkan bahwa mayoritas pengguna Twitter memiliki sentimen positif sebesar 57,6% terhadap ChatGPT, sementara sentimen negatif mencapai 42,4%. Model klasifikasi yang dihasilkan memiliki akurasi sebesar 80%, menunjukkan model klasifikasi yang baik dalam menentukan probabilitas sentimen. Temuan ini memberikan dasar untuk pengembangan teknologi chatbot AI yang lebih baik dan dapat memenuhi kebutuhan pengguna. Hasil penelitian ini memberikan pandangan tentang sentimen pengguna terhadap ChatGPT dan dapat digunakan sebagai acuan bagi pengembangan *chatbot* AI di masa depan..



Kata Kunci: ChatGPT, Analisis Sentimen, *Naive Bayes Classifier*.



KATA PENGANTAR

Segala puji dan syukur kami sampaikan ke hadapan Allah SWT, atas rahmat dan anugerah-Nya, penulis berhasil menyelesaikan karya ilmiah ini dengan judul Implementasi Algoritman *Naïve bayes* untuk analisis sentimen pengguna Twitter terhadap ChatGPT. Penulisan skripsi ini bertujuan untuk memenuhi persyaratan kelulusan dalam mencapai gelar sarjana komputer di Universitas Nusaputra. Dalam kesempatan ini, penulis ingin mengucapkan rasa terima kasih yang tak terhingga kepada:

1. Bapak Dr. Kurniawan, S.T., M.Si, M.M Selaku rektor Universitas Nusaputra Sukabumi.
2. Bapak Anggy Pradiftha J., S.Pd., M.T Selaku Wakil Rektor I Bidang Akademik Universitas Nusa Putra Sukabumi.
3. Bapak Ir. Paikun, S.T., M.T., IPM., Asean eng Selaku Dekan Fakultas Teknik Komputer dan Desain Universitas Nusa Putra Sukabumi
4. Bapak Adhtia Erfina, ST., M.Kom Selaku Kepala Program Studi Sistem Informasi dan juga Dosen Pembimbing I Universitas Nusa Putra Sukabumi.
5. Bapak Mupaat, M.Kom Selaku Dosen Pembimbing II Universitas Nusa Putra Sukabumi.
8. Para Dosen Program Studi Sistem Informasi Universitas Nusa Putra Sukabumi
9. Orang tua, Ayah dan ibu yang senantiasa selalu memberikan dukungan dan doa yang tiada henti untuk keberhasilan saya di masa depan
10. Orang Istimewa, yang selalu memberikan saya dukungan dan motivasi-motivasi yang membangun untuk keberhasilan saya di masa depan. Sehingga membuat saya bersemangat dalam menyelesaikan skripsi
11. Rekan-rekan mahasiswa dan mahasiswi seperjuangan di Program studi sistem informasi universitas Nusa Putra atas bantuan dan kerjasamanya selama Penyusunan Skripsi ini.

Penulis menyadari bahwa skripisi ini masih jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun dari berbagai pihak sangat kami harapkan demi perbaikan. Amin Yaa Rabbal 'Alamiin.

Sukabumi, 03 juli 2023

M Rifki Nurul R A



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik UNIVERSITAS NUSA PUTRA, saya yang bertanda tangan di bawah ini:

Nama : M Rifki Nurul R A

NIM : 20190050042

Program Studi : Sistem Informasi

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty- Free Right*)** atas karya ilmiah saya yang berjudul:

IMPLEMENTASI ALGORITMA KLASIFIKASI *NAIVE BAYES* UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP CHATGPT beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non Eksklusif ini, Universitas Nusa Putra berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Sukabumi

Pada tanggal : 3 Juli 2023


M Rifki Nurul R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN PENULIS.....	ii
PENGESAHAN SKRIPSI	iii
ABSTRACT	iv
ABSTRAK	v
KATA PENGANTAR	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terkait	6
2.2 Landasan Teori	8
2.3 Kerangka berpikir.....	10
BAB III METODE PENELITIAN.....	12
3.1 Tahapan Penelitian	12
3.2 Pengumpulan data	13
3.3 <i>Preprocessing</i>	14
3.3.1 <i>Case Folding</i>	14

3.3.2	<i>Cleaning</i>	15
3.3.3	<i>Tokenizing</i>	15
3.3.4	<i>Text Normalization</i>	15
3.3.5	<i>Stopword Removal</i>	15
3.3.6	<i>Stemming</i>	15
3.3.7	<i>Remove Data Duplicate</i>	15
3.3.8	<i>Remove Data NAN (Not A Number)</i>	15
3.4	Pelabelan Sentimen	16
3.5	Implementasi Klasifikasi Model <i>Naive bayes</i>	17
3.5.1	<i>Split Data</i>	17
3.5.2	Pembobotan kata <i>TF-IDF</i>	18
3.5.3	<i>Cross Validation</i>	18
3.5.4	Klasifikasi Model <i>Naïve Bayes</i>	18
3.5.5	Evaluasi Confusion Matrix	18
3.6	Visualisasi data.....	19
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Pengambilan Data.....	20
4.2	<i>Preporcessing</i>	22
4.2.1	<i>Case Folding</i>	23
4.2.2	<i>Cleaning</i>	23
4.2.3	<i>Tokenizing</i>	24
4.2.4	Text Normalization	25
4.2.5	<i>Stopword removal</i>	26
4.2.6	<i>Stemming</i>	26
4.2.7	<i>Remove Data Duplicate</i>	27
4.2.8	<i>Remove Data NaN</i>	27
4.3	Pelabelan Sentimen	28
4.4	Implementasi Klasifikasi Model <i>Naive bayes</i>	30
4.4.1	<i>Split Data</i>	30
4.4.2	Pembobotan kata <i>TF-IDF</i>	31
4.4.3	<i>Cross Validation</i>	32
4.4.4	Klasifikasi Model <i>Naive Bayes</i>	32
4.4.5	Evaluasi Confusion Matrix	33
4.5	Visualiasi.....	34
BAB V PENUTUP.....		36

5.1	Kesimpulan	36
5.2	Saran.....	36
DAFTAR PUSTAKA		37



DAFTAR GAMBAR

Gambar 1. 1	Grafik data pengguna internet.....	1
Gambar 2. 1	Kerangka berpikir	9
Gambar 3. 1	Metode penelitian.....	12
Gambar 3. 2	Tahapan <i>Scraping</i>	12
Gambar 3. 3	Tahapan <i>Preprocessing</i>	13
Gambar 3. 2	Tahapan <i>Pelabelan Sentimen</i>	15
Gambar 3. 3	Tahapan <i>Algoritma Naive bayes</i>	16
Gambar 4. 1	Script <i>Scraping data</i>	20
Gambar 4. 2	Hasil dari Pengumpulan data	22
Gambar 4. 3	Penampilan dataset dalam file csv	22
Gambar 4. 4	Script <i>Case folding</i>	23
Gambar 4. 5	Script <i>Cleaning</i>	23
Gambar 4. 6	Script <i>Tokenizing</i>	24
Gambar 4. 7	Script <i>Text Normalization</i>	25
Gambar 4. 8	Script <i>Stopword removal</i>	26
Gambar 4. 9	Script <i>Stemming</i>	27
Gambar 4. 10	Script Hapus data duplikat	27
Gambar 4. 11	Script Hapus data NaN.....	28
Gambar 4. 12	Script <i>Pelabelan sentimen</i>	28
Gambar 4. 13	Hasil <i>Pelabelan sentimen</i>	29
Gambar 4. 14	Grafik sentimen positif dan negatif dalam dataset	30
Gambar 4. 15	Script <i>Split data</i>	30
Gambar 4. 23	Script <i>Pembobotan kata TF -IDF</i>	31



Gambar 4. 24 Hasil Pembobotan kata TF -IDF	31
Gambar 4. 25 Script <i>Cross Validation</i>	32
Gambar 4. 26 Script Klasifikasi <i>Naive bayes</i>	32
Gambar 4. 27 Hasil Klasifikasi <i>Naive bayes</i>	33
Gambar 4. 28 Script Evaluasi <i>Confusion Matrix</i>	33
Gambar 4. 29 Hasil Evaluasi <i>Confusion Matrix</i>	33
Gambar 4. 30 Frekuensi kemunculan 10 kata terbanyak	34
Gambar 4. 31 <i>Wordcloud</i> dataset.....	35





DAFTAR TABEL

Tabel 4.1 Sample data hasil Case Folding.....	23
Tabel 4.2 Sample data hasil Cleaning.....	24
Tabel 4.3 Sample data hasil Tokenizing	12
Tabel 4.4 Sample data hasil Text Normalization	25
Tabel 4.5 Sample data hasil Stopword Removal	26
Tabel 4.6 Sample data hasil Stemming.....	27

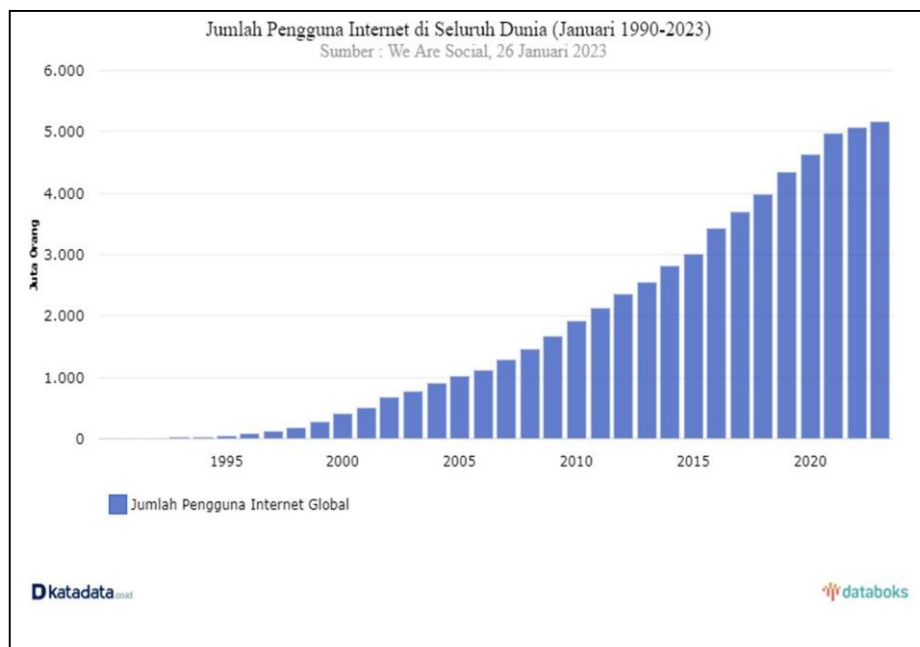




BAB I PENDAHULUAN

1.1 Latar Belakang

Di era sekarang, manusia tidak akan terlepas dari yang namanya teknologi. Teknologi sekarang bertransformasi menjadi kebutuhan primer setiap orang. semuanya tidak terlepas dari yang namanya teknologi seperti E-wallet, M-Banking, E-Learning, E-commerce dan lain sebagainya. Kemajuan ini sebetulnya sangat bermanfaat untuk perkembangan zaman Saat ini, pengguna internet global sudah mencapai 5,15 miliar orang terhitung hingga Januari 2023. Jumlah tersebut mencapai 64,4% dari populasi global yang totalnya 8,01 miliar orang. Jumlah pengguna internet global pada Januari 2023 meningkat 1,9% dibanding periode sama tahun lalu (*year-on-year/yoy*), yang masih 5,01 miliar orang. [1]



Gambar 1.1 Grafik pengguna internet di seluruh dunia

Salah satu teknologi yang saat ini sedang marak di gunakan oleh masyarakat yaitu teknologi *chatbot* AI yang bernama ChatGPT. ChatGPT merupakan singkatan dari *Generative Pre-Trained Transformer* yaitu chatbot berbasis teknologi *Artificial Intelligence* (kecerdasan buatan) yang dapat melakukan interaksi percakapan dengan penggunanya secara canggih. *Chatbot* ini mampu menjawab pertanyaan



user dengan langkah yang sama seperti manusia namun dalam bentuk teks otomatis. [2]. Adapun penulis memanfaatkan ChatGPT untuk membantu penulis ketika melakukan *debugging* dalam proyek mengembangkan aplikasi yang dimana sebelumnya untuk melakukan *debugging* penulis biasanya mengunjungi *stack overflow* yaitu forum konsultasi untuk proyek IT namun sekarang penulis berkonsultasi dengan *chatbot* pintar dari ChatGPT.

Tentunya penulis memilih ChatGPT untuk transformasi dalam hal konsultasi proyek IT dikarenakan banyak keuntungan yang didapatkan yaitu tidak perlu menunggu orang lain menjawab karena ChatGPT dapat menjawab pertanyaan secara langsung dengan jawaban yang akurat. Namun disisi lain penulis juga menduga dampaknya yaitu pekerjaan manusia mungkin akan digantikan oleh teknologi *Artificial Intelligence* (kecerdasan buatan) ini salah satunya ChatGPT, dapat merusak sistem pekerjaan rumah (*homework*) karena pelajar bisa jadi mencontek jawaban langsung dari ChatGPT tanpa berusaha mengerjakan tugas itu secara mandiri dan bahkan terdapat penelitian yang dilakukan oleh peneliti Abu Muna Almaududi Ausat, dkk, berjudul "*Is ChatGPT Dangerous for Lecturer Profession? An In-depth Analysis*", menyimpulkan bahwa ChatGPT dapat memberikan bantuan bagi para dosen dalam beberapa aspek, seperti menghasilkan teks awal dan mempercepat proses penulisan. Namun, pada saat yang sama, ChatGPT juga dapat menimbulkan tantangan dan risiko tertentu. Oleh karena itu, dosen perlu menjaga nilai-nilai inti dari profesi mereka dan tetap aktif dalam memerangi tindakan kecurangan akademik. Hal ini penting mengingat teknologi seperti ChatGPT dapat digunakan dalam kecurangan akademik.[3] Karena terdapat dugaan tersebut, penulis merasa diperlukan untuk adanya analisis sentimen pengguna twitter terhadap chatGPT. Dengan begitu dapat diketahui mengenai pro dan kontra terhadap chatGPT serta memberikan manfaat yang dapat dijadikan sebagai informasi untuk masyarakat dalam penggunaan ChatGpt.

Dalam penelitian ini, digunakan algoritma klasifikasi Naïve Bayes dengan subjek penelitian berupa pengguna Twitter yang melakukan unggahan (tweet) yang mengandung kata-kata chatGPT. Algoritma Naïve Bayes dipilih karena telah terbukti kehandalannya oleh peneliti-peneliti sebelumnya, salah satu di antaranya yaitu peneliti Dinda Tasya Mahardika. mengenai "analisis sentimen masyarakat

Indonesia terhadap *non Fungible token* (nft) menggunakan algoritma *naive bayes* Classifier Penelitian ini bertujuan menganalisis respon masyarakat terhadap konten di YouTube melalui analisis sentimen dari komentar-komentar yang ada. Metode yang digunakan adalah TF-IDF dan Klasifikasi Naïve Bayes. Hasil penelitian menunjukkan akurasi sebesar 64% dengan precision kelas positif sebesar 63% dan recall 99%, precision kelas netral sebesar 83% dengan recall 0.7%, sementara untuk kelas negatif precision 0% dan recall 0%. Perlu lebih memperhatikan klasifikasi kelas negatif untuk mendapatkan hasil yang lebih seimbang. [4]

Berdasarkan permasalahan yang ada, ini menjadi latar belakang penulis untuk mengambil judul “IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP CHATGPT”.

1.2 Rumusan Masalah

Dengan melihat latar belakang masalah sebagaimana yang telah dijelaskan sebelumnya, maka yang menjadi pokok masalah adalah

1. Bagaimana tanggapan pengguna twitter terhadap kehadiran ChatGpt
2. Bagaimana kinerja algoritma *Naive Bayes* dalam menganalisis sentimen pengguna Twitter terhadap ChatGPT dalam dataset yang digunakan
3. Seberapa akurat hasil sentimen analisis ChatGPT ini dengan memakai metode Algoritma *Naive Bayes* dengan bahasa pemrograman *Python*

1.3 Batasan Masalah

Dengan melihat latar belakang masalah sebagaimana yang telah dijelaskan sebelumnya, maka yang menjadi pokok masalah adalah

1. Fokus pada analisis sentimen pengguna Twitter terhadap ChatGPT: Penelitian ini akan difokuskan pada analisis sentimen yang berkaitan dengan ChatGPT dan pengguna Twitter yang membicarakannya.
2. Menggunakan metode analisis sentimen dengan algoritma Naive Bayes: Penelitian ini akan menggunakan algoritma Naive Bayes untuk menganalisis sentimen pengguna Twitter terhadap ChatGPT.

3. Penelitian ini akan difokuskan pada analisis sentimen dalam bahasa Inggris karena pengguna Twitter yang membicarakan ChatGPT dapat berasal dari berbagai negara.
4. Fokus pada sentimen positif, negatif, dan netral: Analisis sentimen yang dilakukan akan difokuskan pada sentimen positif, negatif, dan netral terhadap ChatGPT yang diungkapkan oleh pengguna Twitter dalam dataset yang digunakan.

1.4 Tujuan Penelitian

1. Penelitian ini bertujuan untuk mengetahui bagaimana respon masyarakat terhadap ChatGPT dengan Mengetahui respon masyarakat terhadap ChatGPT dengan menerapkan algoritma Naive Bayes dan bahasa pemrograman Python.
2. Mengukur kinerja algoritma Naive Bayes dalam menganalisis sentimen pengguna Twitter terhadap ChatGPT menggunakan dataset yang relevan dengan topik ChatGPT.
3. Mencari wawasan baru melalui visualisasi data bagaimana tanggapan masyarakat mengenai ChatGPT

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Bagi Penulis
 - a. Meningkatkan pemahaman dan wawasan tentang analisis sentimen menggunakan metode Klasifikasi Naïve Bayes.
 - b. Memperdalam bahasa pemrograman python dalam hal data science
2. Bagi masyarakat
 - a. Memberikan informasi tentang hasil dari analisis sentimen mengenai kehadiran ChatGPT
3. Bagi perguruan tinggi
 - a. Dalam penelitian yang serupa diharapkan dapat digunakan sebagai bahan referensi mahasiswa.

1.6 Sistematika Penulisan

Sistematika penulisan dalam Skripsi terdiri dari lima (5) bab, yaitu:

BAB I: PENDAHULUAN

Pada bab ini, dijelaskan mengenai Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II: TINJAUAN PUSTAKA

Pada bab ini, dijelaskan mengenai Penelitian Terkait, Landasan Teori, dan Kerangka Berpikir.

BAB III: METODOLOGI PENELITIAN

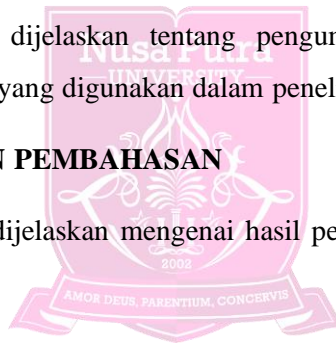
Pada bab ini, dijelaskan tentang pengumpulan data, tahapan proses penelitian, dan metode yang digunakan dalam penelitian.

BAB IV: HASIL DAN PEMBAHASAN

Pada bab ini, dijelaskan mengenai hasil pengujian data dan pembahasan hasil penelitian.

BAB V: PENUTUP

Pada bab ini, dijelaskan mengenai kesimpulan dan saran yang diambil dari penelitian ini.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian yang dilakukan Dinda Tasya Mahardika. mengenai “Analisis sentimen masyarakat indonesia terhadap *non Fungible token* (nft) menggunakan algoritma *Naive bayes Classifier* Penelitian ini bertujuan menganalisis respon masyarakat terhadap konten di YouTube melalui analisis sentimen dari komentar-komentar yang ada. Metode yang digunakan adalah TF-IDF dan Klasifikasi Naïve Bayes. Hasil penelitian menunjukkan akurasi sebesar 64% dengan precision kelas positif sebesar 63% dan recall 99%, precision kelas netral sebesar 83% dengan recall 0.7%, sementara untuk kelas negatif precision 0% dan recall 0%. Perlu lebih memperhatikan klasifikasi kelas negatif untuk mendapatkan hasil yang lebih seimbang. [4]

Penelitian yang dilakukan oleh tim peneliti yang dipimpin oleh Diah Rahmalenia, membahas tentang "Analisis persepsi pemulihan ekonomi di Indonesia setelah masa pandemi COVID-19 pada platform Twitter menggunakan Algoritma Klasifikasi Naïve Bayes". Hasil dari penelitian tersebut menunjukkan bahwa Algoritma Klasifikasi Naïve Bayes mampu mengklasifikasikan data tweet dengan tingkat akurasi mencapai 78%. Hasil klasifikasi menunjukkan bahwa precision prediksi positif mencapai 96%, sedangkan prediksi negatif mencapai 31%. Selain itu, recall yang diperoleh dari true positif adalah sebesar 78%, sementara true negatif mencapai 75%. Hasil penelitian ini menunjukkan bahwa penggunaan Algoritma Klasifikasi Naïve Bayes dalam melakukan analisis sentimen pada tweet publik telah mencapai tingkat performa yang diharapkan.[5].

Penelitian yang dilakukan oleh tim peneliti yang dipimpin oleh Anisa Sholihat, membahas tentang "Analisis Sentimen Terhadap Investasi Ilegal pada Twitter Menggunakan Algoritma Naive Bayes". Penelitian ini menggunakan metode Naïve Bayes dengan alat bantu Orange. Pengumpulan data dilakukan pada bulan Maret 2022. Hasil penelitian menunjukkan terdapat 152 data sentimen positif, 244 data sentimen negatif, dan 175 data sentimen netral, dengan tingkat akurasi mencapai 81,069%. Hal ini menandakan bahwa tingkat sentimen negatif lebih dominan dibandingkan dengan sentimen positif. Dengan adanya penelitian ini, diharapkan



masyarakat lebih berhati-hati dalam memilih investasi, terutama untuk menghindari investasi ilegal.[6]

Penelitian yang dilakukan oleh Yohakim Benedictus Samponu dan Kursini membahas tentang "Optimasi Algoritma Naive Bayes dengan Metode Cross Validation untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan tepat waktu". Penelitian ini bertujuan untuk memprediksi tingkat kelulusan tepat waktu dengan menggunakan analisis menggunakan teknik klasifikasi data mining.

Algoritma Naïve Bayes digunakan sebagai model dalam penelitian ini dan dijadikan acuan untuk analisis. Penulis melakukan berbagai skenario percobaan dengan teknik cross validation untuk membandingkan hasil dan menilai tingkat akurasi yang dihasilkan dari penelitian ini.

Hasil penelitian menunjukkan bahwa dengan penambahan skenario pengujian Cross Validation, terdapat peningkatan akurasi sebesar 2% dari hasil pengujian yang dilakukan. Hal ini menunjukkan bahwa penggunaan Cross Validation dalam penelitian dapat meningkatkan keakuratan prediksi tingkat kelulusan tepat waktu secara signifikan. Penemuan ini diharapkan dapat memberikan manfaat dalam meningkatkan efisiensi dan ketepatan dalam merencanakan program pendidikan dan mengidentifikasi faktor-faktor yang berpengaruh terhadap kelulusan tepat waktu.[7].

Studi yang dilakukan oleh Firman Tempola dan rekan-rekannya membahas tentang "Perbandingan Klasifikasi antara K-NN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan Metode K-Fold Cross Validation". Penelitian ini bertujuan untuk melakukan klasifikasi aktivitas gunung berapi menggunakan faktor kegempaan. Hasil penelitian menunjukkan bahwa metode Naive Bayes memiliki akurasi yang lebih tinggi dan lebih stabil dibandingkan dengan metode K-NN. Penggunaan Naive Bayes dapat dianggap sebagai pilihan yang lebih baik dalam klasifikasi ini. [8]

Penelitian yang dilakukan oleh Mubin UI Haque, dkk yang berjudul "‘I think this is the most disruptive technology’ Exploring Sentiments of ChatGPT Early Adopters using Twitter Data" Penelitian ini dilakukan pada tahun 2022 dengan mengambil populasi data dari user tweet pengguna awal ChatGPT yang berjumlah 10.732 metode yang digunakan adalah Mixed Methods pada penelitian ini

bertujuan untuk mengetahui potensi keberhasilan atau kegagalan teknologi, serta kekuatan dan kelemahannya. Dari ChatGPT. Hasil penelitian kami menunjukkan bahwa mayoritas pengguna awal memiliki mengungkapkan sentimen yang sangat positif terkait dengan topik seperti Gangguan terhadap pengembangan perangkat lunak, Hiburan dan melatih kreativitas. Hanya sebagian kecil pengguna yang menyatakan kekhawatiran tentang isu-isu seperti potensi penyalahgunaan ChatGPT, terutama terkait topik-topik seperti Dampak terhadap aspek pendidikan.[9]

2.2 Landasan Teori

2.2.1. ChatGPT

ChatGPT adalah model pemrosesan bahasa alami tingkat lanjut. Model dilatih sebelumnya pada kumpulan data besar, memungkinkannya membuat bahasa yang koheren, sesuai konteks, dan sering tidak dapat dibedakan dari barang-barang yang ditulis manusia. Itu menggunakan arsitektur transformator yang menggabungkan manfaat model generatif dengan skalabilitas. Itu menciptakan bahasa alami menggunakan mekanisme perhatian diri, yang memungkinkannya mempelajari konteks dialog dan penawaran balasan yang lebih akurat[10]. ChatGPT memungkinkan kita dapat bertanya dan langsung mendapatkan jawaban nya dengan baik.

2.2.2. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari Pengolahan Bahasa Alami (Natural Language Processing/NLP) yang bertujuan untuk mengembangkan sistem yang dapat mengenali dan mengekstraksi pendapat atau opini dalam bentuk teks. Banyak informasi berbentuk teks yang tersebar di internet, seperti forum, blog, media sosial, serta situs yang berisi ulasan dan review. Dengan menggunakan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur dan dapat dipahami secara lebih mudah dan efisien. Hal ini sangat berguna untuk mengumpulkan dan menganalisis opini atau tanggapan masyarakat terhadap suatu topik atau produk tertentu, dan dapat digunakan dalam berbagai bidang,

seperti pemasaran, penelitian pasar, dan analisis brand..[11] Analisis sentimen telah banyak digunakan dalam berbagai aplikasi, termasuk di media sosial seperti Twitter.

2.2.3. Algoritma Klasifikasi *Naive bayes*

algoritma *Naive bayes* adalah sebuah metode klasifikasi yang berakar pada teorema Bayes dengan asumsi yang sangat kuat (*naive*) akan independensi antara fitur-fiturnya. Algoritma ini populer digunakan dalam aplikasi klasifikasi teks seperti pengenalan spam email dan klasifikasi dokumen. Algoritma ini memerlukan dataset pelatihan untuk melakukan pelatihan model, dan kemudian model tersebut digunakan untuk melakukan klasifikasi pada data yang tidak dilihat sebelumnya.[12] Selain itu, algoritma *Naive Bayes* juga sangat efisien dalam pengolahan data yang besar, sehingga dapat diaplikasikan pada data dengan jumlah yang sangat besar, seperti data dari media sosial yang seringkali mencapai jutaan data setiap harinya. Hal ini memungkinkan *algoritma Naive bayes* digunakan dalam berbagai aplikasi, seperti klasifikasi dokumen, klasifikasi email spam, analisis sentimen, dan lain sebagainya. Oleh karena itu, algoritma *Naive bayes* menjadi salah satu algoritma yang banyak digunakan dalam pengolahan teks dan data besar saat ini. Rumus algoritma *Naive bayes* dapat dinyatakan sebagai berikut:

$$p(C_k | \mathbf{x}) = \frac{p(C_k) p(\mathbf{x} | C_k)}{p(\mathbf{x})}$$

Di mana:

$P(C | X)$ adalah probabilitas kelas C jika fitur X terlihat

$P(X | C)$ adalah probabilitas fitur X jika kelas C benar

$P(C)$ adalah probabilitas kelas C

$P(X)$ adalah probabilitas fitur X

Algoritma *Naive bayes* menghitung probabilitas $P(C | X)$ dari kelas C dengan mengalikan probabilitas $P(X | C)$ dari setiap fitur X dengan probabilitas $P(C)$ dari kelas C, dan kemudian membagi hasilnya dengan probabilitas $P(X)$ dari fitur X. Algoritma ini dianggap "*naive*" karena mengasumsikan bahwa

setiap fitur adalah independen satu sama lain, meskipun pada kenyataannya fitur-fitur tersebut mungkin saling terkait. Namun, meskipun asumsi tersebut sederhana, *Naive bayes* terbukti sangat efektif dalam banyak kasus, termasuk dalam analisis sentimen.

2.2.4. Python

Python adalah bahasa pemrograman serbaguna yang bisa dijalankan pada hampir semua arsitektur sistem, dan bisa digunakan untuk berbagai aplikasi di banyak bidang, mulai dari web development hingga machine learning.[13] *Python* menyediakan berbagai library dan tools yang memudahkan implementasi analisis sentimen menggunakan algoritma *Naive bayes*. *Natural Language Toolkit* (NLTK): NLTK adalah library *Python* yang digunakan untuk memproses bahasa alami. NLTK menyediakan berbagai fitur seperti tokenisasi, stemming, dan pembentukan kosa kata yang berguna untuk tahapan pre-processing dalam analisis sentimen.

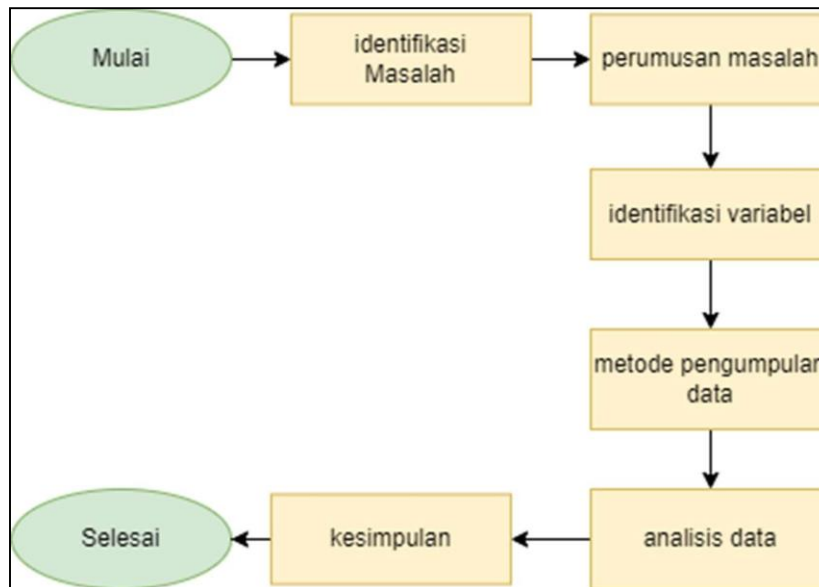
2.2.5. Web Scraping

Web scraping adalah sebuah cara pengambilan suatu data atau informasi tertentu dengan jumlah besar untuk nantinya digunakan dalam berbagai keperluan seperti riset, analisis dan lainnya. Dibandingkan dengan melakukan survei secara manual.[14] dalam penelitian ini sumber data yang akan di kumpulkan berasal dari platform media sosial yaitu twitter . Teknik ini memungkinkan pengguna untuk mengambil data dari Twitter seperti tweet, metadata *user*, *retweet*, *like*, *hashtag*, dan banyak lagi.

2.2.6. Visualisasi

Visualisasi data (*data visualization*) merupakan rangkaian proses yang akan dilakukan setiap data analis untuk menampilkan data atau informasi dalam bentuk yang agar mudah dipahami oleh orang awam, seperti grafik, angka dan lain sebagainya.[15]

2.3 Kerangka berpikir



Gambar 2.1 Kerangka berpikir

Penelitian ini bermula dari identifikasi masalah terkait dengan dampak kehadiran teknologi ChatGPT pada masyarakat. Masalah ini kemudian dirumuskan menjadi pertanyaan penelitian yang lebih spesifik, yaitu "Bagaimana sentimen masyarakat terhadap teknologi ChatGPT?". Setelah merumuskan pertanyaan penelitian, dilakukan identifikasi variabel untuk memperjelas hubungan antara variabel dependen dan variabel independen. Variabel dependen dalam penelitian ini adalah opini atau sentimen masyarakat terhadap teknologi ChatGPT yang dapat bersifat positif atau negatif. Sementara itu, variabel independen adalah data tweet yang mengandung kata kunci #ChatGpt.

Setelah melakukan identifikasi variabel, dilakukan metode pengumpulan data yang sesuai dengan tujuan penelitian. Dalam hal ini, data tweet yang mengandung kata kunci #ChatGpt dikumpulkan dan dianalisis. Metode analisis data yang digunakan dalam penelitian ini adalah analisis sentimen *Naive Bayes*. Analisis sentimen *Naive Bayes* dipilih karena metode ini dapat memberikan hasil yang akurat dalam mengklasifikasikan sentimen positif dan negatif dari data tweet yang dikumpulkan. Setelah melakukan analisis data, dilakukan penyusunan kesimpulan berdasarkan hasil penelitian. Kesimpulan tersebut harus memberikan jawaban yang jelas dan akurat terhadap pertanyaan penelitian yang telah dirumuskan sebelumnya.

BAB V PENUTUP

5.1 Kesimpulan

1. Penerapan algoritma *Naive Bayes* untuk mengklasifikasikan sentimen positif dan negatif pada data tweet terhadap ChatGPT menghasilkan nilai akurasi sebesar 80%, yang berarti model yang dibangun sudah mampu merepresentasikan klasifikasi dengan baik.
2. Informasi yang didapatkan dari analisis sentimen dan klasifikasi pada penelitian ini yaitu sentimen positif sebesar 57,6% dan sentimen negatif sebesar 42,4% dari presentase tersebut menunjukkan bahwa dengan adanya ChatGPT pengguna twitter lebih banyak bersentimen positif
3. Informasi yang diperoleh dari visualisasi grafik 10 kata terbanyak yang diucapkan dan word cloud menunjukkan bahwa kata "ChatGPT" paling banyak digunakan, diikuti oleh kata-kata seperti "use", "like", "write", "make", "OpenAI", "well", "think", "time", dan "good" Untuk membuktikan kegunaan ChatGPT penulis meminta ChatGPT untuk mengkombinasikan 10 kata tersebut menjadi 1 kalimat yang utuh dan jawaban ChatGPT itu adalah *“ChatGPT is a well-known language model created by OpenAI that is used for various tasks, like write and generating content, is make good tool, good performance, and ability to think at any time.”* Artinya bahwa ChatGPT, yang dikenal sebagai model bahasa canggih yang diciptakan oleh OpenAI, dapat digunakan untuk berbagai tugas seperti menulis dan menghasilkan konten, sehingga menjadikannya alat yang bagus dengan performa yang baik dan kemampuan berpikir di setiap saat.

5.2 Saran

Untuk peneliti selanjutnya yang ingin melakukan penelitian terkait dengan ChatGPT, disarankan untuk menyertakan pengkajian mengenai kelemahan ChatGPT dan merangkumnya dalam kuesioner yang dapat diisi oleh masyarakat. Hal ini akan membantu untuk melihat apakah masyarakat masih memiliki sentimen positif terhadap ChatGPT meskipun kelemahannya telah diungkap. silahkan mencoba untuk melaksanakan penelitian tersebut.



DAFTAR PUSTAKA

- [1] Annur Cindy Mutia, "Jumlah Pengguna Internet Global Tembus 5,16 Miliar Orang pada Januari 2023," *databooks*, Feb. 03, 2023. <https://databoks.katadata.co.id/datapublish/2023/02/03/jumlah-pengguna-internet-global-tembus-516-miliar-orang-pada-januari-2023> (accessed Feb. 19, 2023).
- [2] Suud Hefty, "Apa Itu Chat GPT? Ini Arti Kata, Keunggulan dan Cara Menggunakannya, Chatbot Viral di Media Sosial," *TribunJatim.com*, Feb. 04, 2023. <https://jatim.tribunnews.com/2023/02/04/apa-itu-chat-gpt-ini-arti-kata-keunggulan-dan-cara-menggunakannya-chatbot-viral-di-media-sosial> (accessed Feb. 10, 2023).
- [3] A. M. Almaududi Ausat, Suherlan, and H. K. Azzaakiyyah, "Is ChatGPT Dangerous for Lecturer Profession? An In-depth Analysis."
- [4] D. Tasya Mahardika, "ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP NON FUNGIBLE TOKEN (NFT) MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER SKRIPSI," 2022.
- [5] D. Rahmalenia, D. Maulinda, and F. Nur Ocktafiani, "ANALISIS SENTIMEN PEMULIHAN EKONOMI DI INDONESIA PASCA MASA PANDEMI COVID-19 PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER," 2022.
- [6] A. Sholihat, F. Bei, and R. Ainaya, "TWITTER TWEET: ANALISIS SENTIMEN INVESTASI ILEGAL MENGGUNAKAN ALGORITMA NAIVE BAYES," 2022.
- [7] Y. Benedictus Samponu, "OPTIMASI ALGORITMA NAIVE BAYES MENGGUNAKAN METODE CROSS VALIDATION UNTUK MENINGKATKAN AKURASI PREDIKSI TINGKAT KELULUSAN TEPAT WAKTU," *Print) Jurnal ELTIKOM*, vol. 1, no. 2, 2017.
- [8] F. Tempola, M. Muhammad, and A. Khairan, "PERBANDINGAN KLASIFIKASI ANTARA KNN DAN NAIVE BAYES PADA PENENTUAN STATUS GUNUNG BERAPI DENGAN K-FOLD CROSS VALIDATION COMPARISON OF CLASSIFICATION BETWEEN KNN AND NAIVE BAYES AT THE DETERMINATION OF THE VOLCANIC STATUS WITH K-FOLD CROSS VALIDATION," vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik20185983.
- [9] M. U. Haque, I. Dharmadasa, Z. T. Sworna, R. N. Rajapakse, and H. Ahmad, "'I think this is the most disruptive technology': Exploring Sentiments of ChatGPT Early Adopters using Twitter Data," Dec. 2022, [Online]. Available: <http://arxiv.org/abs/2212.05856>
- [10] Ö. Aydin and E. Karaarslan, "Is ChatGPT Leading Generative AI? What is Beyond Expectations?" [Online]. Available: <https://twitter.com/aaronsiim>
- [11] Adminlp2m, "Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya," *LP2M*, Feb. 21, 2022. <https://lp2m.uma.ac.id/2022/02/21/analisis->

- sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/ (accessed Feb. 25, 2023).
- [12] "Naive Bayes classifier - Wikipedia."
https://en.wikipedia.org/wiki/Naive_Bayes_classifier (accessed Feb. 25, 2023).
- [13] Faradila A, "Apa Itu Python? Yuk Kenalan dengan Bahasa Pemrograman Ini!," Nov. 23, 2022. <https://www.hostinger.co.id/tutorial/python-adalah> (accessed Feb. 25, 2023).
- [14] Meilinaeka, "Web Scraping : Pengertian dan Fungsinya dalam Pengambilan Data | Direktorat Pusat Teknologi Informasi," *Blogs*, Dec. 24, 2022.
<https://it.telkomuniversity.ac.id/web-scraping-pengertian-dan-fungsinya-dalam-pengambilan-data/> (accessed Mar. 17, 2023).
- [15] Putri Anugrah, "Apa Itu Data Visualization dalam Data Science | Danacita," *Blog*, Oct. 12, 2022. <https://danacita.co.id/blog/apa-itu-data-visualization-dalam-data-science/> (accessed Mar. 17, 2023).
- [16] JustAnotherArchivist, "A social networking service scraper in Python," *Github*, 2018. <https://github.com/JustAnotherArchivist/snsrape> (accessed Mar. 18, 2023).
- [17] Wibowo Antoni, "10 Fold-Cross Validation – MTI," *Binus Articles*, 2017.
<https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/> (accessed Apr. 17, 2023).
- [18] Administrator, "Apa itu Word Cloud? Kenali Fungsinya untuk Aplikasi Digital Monitoring," Jan. 24, 2023. <https://ivosights.com/read/artikel/word-cloud-apa-itu-kenali-fungsinya-untuk-aplikasi-digital-monitoring> (accessed May 01, 2023).

Lampiran 1 Curriculum Vitae***CURRICULUM VITAE***

Nama : M Rifki Nurul R A

Nim 20190050042

Tempat, Tanggal Lahir : Sukabumi, 30 November 2022

Alamat : Kp Cicatih RT/RW 005/001

Kecamatan

Kabupaten : Sukabumi

Pendidikan



1. SDN 1 NYANGKOWEK

2. MTS ASSA'ADAH

3. SMK INFORMATIKA CBI

Pekerjaan : Mahasiswa

Publikasi / Karya Ilmiah : Implementasi Algoritma Klasifikasi Naive Bayes
Untuk Analisis Sentimen Pengguna Twitter
Terhadap ChatGPT

