Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: journaljoecy@gmail.com

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X **(Printed)**
5 (2) 2025, Page: 3164-3176

Perbandingan Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Terhadap Analisis Sentimen *Review* Pengguna Aplikasi Video Pendek

Saromah[1], Anjelis Ratuwulan[2], Samroh[3], Anisa Yuliantin[4]
[1,2] Program Studi Teknik Informatika, [3,4] Program Studi Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Mercusuar, Bekasi
Corresponding Author: Saromah, Email: **Saromah@mercusuar.ac.id**

Published: May,2025

ABSTRAK

Perkembangan teknologi informasi dan jejaring sosial telah menghasilkan ledakan konten digital. Aplikasi video pendek telah menjadi fenomena sosial yang signifikan dalam era digital, bagi calon pengguna, penyedia konten, pemasar menjadi masalah yang krusial dalam mengukur efektivitas kampanye pemasaran untuk mencapai audiens yang lebih luas maka dibutuhkan pemilihan aplikasi yang tepat untuk mengukur keberhasilan kampanye pemasaran melalui review pengguna aplikasi video pendek melalui *Google Play Store*. Penelitian ini dilakukan dengan menggunakan perbandingan antara Algoritma K-Nearest Neighbor dan Naïve bayes dapat digunakan untuk mengklasifikasi ulasan menjadi kategori sentimen yang berbeda seperti positif, negatif atau netral. Untuk melihat tingkat akurasi antara 2 Algoritma dalam penelitian ini menggunakan Google Colab sebagai tools pengambilan data dan rapidminer sebagai tools preprocessing data. Proses analisis sentimen terhadap ulasan pengguna aplikasi video pendek TikTok dengan menggunakan metode *K-NN* dan *Naïve Bayes*. Data awal yang diperoleh melalui proses web crawling dari *Google Play Store* mencakup1000 ulasan, yang kemudian mengalami proses *cleaning* sehingga tersisa 998 ulasan yangbersih dansiap digunakan. Dalam uji validasi, metode *K- NN* menghasilkan akurasi sekitar92,48%, dengan *Precision* positif sebesar 92,77% dan *Precision* negatif sebesar 75,00%, serta *Recall* positif sebesar 99,56% dan *Recall* negatif sebesar 14,45%. Sebaliknya, metode *Naïve Bayes* mencapai akurasi sekitar 99,50%, dengan *Precision*positif sebesar 100,00% dan Precision negatif sebesar 94,32%, serta *Recall* positif sebesar99,45% dan Recall negatif sempurna (100,00%). Hasil ini menunjukkan bahwa metode Naïve Bayes memberikan kinerja yang lebih baik dibandingkan dengan metode *K-NN* dalam menganalisis sentimen ulasan TikTok, dengan akurasi yang tinggi dan kemampuan yang sangat baik dalam mengklasifikasikan ulasan positifdan negatif. Oleh karena itu, metode *Naïve Bayes* disarankan untuk digunakan dalam analisis sentimen ulasan TikTok dalam konteks penelitian ini

Kata kunci: ***Analisis sentiment, K-Nearest Neighbor, Naïve Bayes***

ABSTRACT

The development of information technology and social networks has resulted in an explosion of digital content. Short video applications have become a significant social phenomenon in the digital era, for potential users, content providers, marketers it is a crucial problem in measuring the effectiveness of marketing campaigns to reach a wider audience, so selecting the right application is needed to measure the success of marketing campaigns through user reviews. short video application via Google Play Store. This research was conducted using a comparison between the K-Nearest Neighbor and Naïve Bayes algorithms which can be used to classify reviews into different sentiment categories such as positive, negative or neutral. To see the level of accuracy between the 2 algorithms in this research, Google Colab was used as a data retrieval tool and Rapidminer as a data preprocessing tool. The process of sentiment analysis of user reviews of the short video application TikTok using the K-NN and Naïve Bayes methods. Initial data obtained through the web crawling process from the Google Play Store included 1000 reviews, which then underwent a cleaning process so that 998 reviews were left that were clean and ready to use. In the validation test, the K-NN method produced an accuracy of around 92.48%, with a positive precision of 92.77% and a negative precision of 75.00%, as well as a positive recall of 99.56% and a negative recall of 14.45%. In contrast, the Naïve Bayes method achieves an accuracy of around 99.50%, with a positive Precision of 100.00% and a negative Precision of 94.32%, as well as a positive Recall of 99.45% and a perfect negative Recall (100.00%). These results show that the Naïve Bayes method provides better performance compared to the K-NN method in analyzing the sentiment of TikTok reviews, with high accuracy and excellent ability in classifying positive and negative reviews. Therefore, the Naïve Bayes method is recommended for use in sentiment analysis of TikTok reviews in the context of this research

Keywords: Sentiment analysis, K-Nearest Neighbor, Naïve Bayes

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Atibution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

PENDAHULUAN

Dalam beberapa tahun terakhir, platform self-media video pendek menghasilkan nilai komersial yang tinggi (Wati, 2020:20). Menurut data pada "*we are sosial*" statistik pada bulan Januari 2023, aktivitas harian platform video pendek itu mencapai lebih dari 1,05 miliar, dan volume interaksi video pujian dan komentar telah mencapai 100 miliar. Pertumbuhan konten video sendiri dapat dengan mudah diunggah ke internet hal ini di dukung dengan kemajuan kecepatan internet. Banyak platform maupun aplikasi yang menyediakan dukungan pembuatan video dan dengan hal menarik pada pengguna smartphone.

Perkembangan teknologi informasi dan jejaring sosial telah menghasilkan memahami respon pengguna secara holistik. Google Play adalah layanan konten digital milik Google yang melingkupi produk-produk seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media. Layanan produk yang dapat ditemui di Google Play adalah aplikasi video pendek yang sedang terpopuler yaitu, Tiktok. Layanan ini dapat diakses baik melalui web, aplikasi android (Play Store), dan Google. Dalam Google Play, selain toko produk-produk online juga dilengkapi adanya fitur penilaian bagi pelanggan untuk memberikan ulasan tentang kelebihan ataupun kekurangan dalam masing masing penggunaan aplikasi video pendek Ulasan pengguna bisa bermacam-macam bentuknya, mulai dari kalimat yang halus sampai kalimat ledakan konten digital, terutama dalam yang kasar, selain memberikan ulasan bentuk video sedang tren. Aplikasi video pendek seperti TikTok telah menjadi fenomena global dengan jutaan pengguna aktif setiap harinya. Pengguna dapat memilih dari berbagai filter, efek visual, dan fitur editing lainnya untuk membuat video yang kreatif dan menarik Pengguna aplikasi ini tidak hanya menonton video, tetapi juga memberikan ulasan atau komentar terkait konten yang mereka saksikan. Dalam ekosistem aplikasi video pendek, ulasan pengguna memiliki nilai informatif yang signifikan. Ulasan tersebut mencerminkan pandangan, opini, dan perasaan pengguna terhadap video yang mereka tonton. Dalam konteks ini, analisis sentimen, yaitu proses mengklasifikasikan ulasan ke dalam kategori sentimen positif, negatif, atau netral, menjadi penting untuk para pengguna juga memberikan skor bintang (antara 1-5) tergantung penilaian yang diberikan oleh pengguna tersebut dan terakhir pengguna juga bisa memberikan skor "like" terhadap komentar seseorang yangmewakili kepuasan atau kekecewaan seseorang terhadap aplikasi video Pendek tersebut.Ulasan dari pengguna digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Data hasil dari penelitian dapat diolah dan digunakan oleh perusahaan, sehingga perusahaan dapat memahami sentimen apa yang mendominasi pemikiran konsumen tentang perusahaan atau aplikasi mereka, memahami konsumen atau pelanggan merupakan cara agar perusahaan dapat terus melihat *feedback* maupun jasa yang diberikan dari konsumen terhadap aplikasi untuk perusahaan. Permasalahan yang berkaitan dengan beberapa ulasan tentang aplikasi video pendek di*Google Play Store* fokus utamanya adalah *review* pengguna aplikasi video pendek yangada di situs *Google Play Store*. Ulasan pengguna aplikasi video pendek bisa dipengaruhioleh beberapa hal yang belum menjadi perhatian baik dari pihak

pengembang aplikasi . Hal ini mungkin terjadi karena adanya beberapa faktor yang harus diperbaiki dan belumdiketahui oleh pihak Aplikasi. Berdasarkan dari penjelasan di atas diperlukan sebuah metode untuk dapat mengklasifikasikan dari suatu ulasan tersebut secara otomatis, apakah positif atau negatif yang nantinya akan menghasilkan suatu akurasi dari pengolahan data ulasan dalam aplikasi video pendek terpopuler.

Pada penelitian ini akan menganalisis aplikasi video pendek terpopuler yang ada di *Google play store* (Tiktok). Karena aplikasi-aplikasi tersebut merupakan aplikasi yang memiliki banyak penilaian di *Google Play Store* sebagai pusat unduhan aplikasi. Penilaian kepuasan pengguna terhadap aplikasi video pendek dinilai sangat penting karena dapat mengukur keberhasilan aplikasi dan sebagai bahan evaluasi perusahaan dalam meningkatkan layanan aplikasi dan bermanfaat bagi penyedia konten dan pemasar untuk mengukur efektivitas kampanye pemasaran,mengevaluasi respon audiens terhadap konten yang mereka bagikan dan membuat keputusan yang lebih cerdas dalam pengembangan konten dan startegi pemasaran dan juga dapat membantu calon pengguna dalam memilih aplikasi video pendek yang paling sesuai dengan preferensi mereka berdasarkan ulasan sentimen pengguna lainnya.

Hal tersebut di mungkinkan dengan analisa sentimen. Analisis sentimen atau yang sering disebut dengan opini mining merupakan sebuah proses menemukan gagasan seseorang mengenai beberapa ide atau pokok yang disampaikan pengguna tersebut, opini yang dimuat di media sosial jumlahnya terlalu banyak untuk diproses secara manual. Oleh sebab itu penelitian ini ingin mengetahui sejauh mana analisis review pengguna aplikasi video pendek pada aplikasi setelah menggunakan aplikasi tersebut. Berdasarkan pada uraian latar belakang yang telah dipaparkan maka dalam penelitian ini mengajukan topik dengan judul " Perbandingan Algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes terhadap Analisis sentimen Review pengguna Aplikasi video pendek "analisis sentimen terhadap pengguna aplikasi video pendek dimulai dari tahap preprocessing sampai tahap analisa sentimen dengan menggunakan metode data mining dia antaranya K- Nearest Neighbor dan Naïve Bayes dengan bagaimana mengukur kualitas hasil analisis dari masing-masing algoritma klasifikasi ulasan aplikasi video pendek.

LANDASAN TEORI

Perbandingan

Menurut Marhiyanto (2000:57) perbandingan berasal dari kata banding yang berarti persamaaan, selanjutnya membandingkan mempunyai arti mengadu dua hal untuk diketahui perbandinganya. Perbandingan diartikan sebagai selisih persamaan.

Menurut Basah (1994:7) perbandingan merupakan suatu metode pengkajian atau penyelidikan dengan mengadakan perbandingan di antara dua objek kajian atau lebih untuk menambah dan memperdalam pengetahuan tentang objek yang dikaji. Jadi di dalam perbandingan ini terdapat objek

yang hendak diperbandingkan yang sudah diketahui sebelumnya, akan tetapi pengetahuan ini belum tegas dan jelas.

Algoritma

Algoritma adalah langkah-langkah terstruktur dan terdefinisi dengan jelas yang digunakan untuk menyelesaikan masalah atau mencapai tujuan tertentu. Dalam dunia komputasi dan pemrograman, algoritma adalah urutan instruksi yang terorganisasi dengan baik yang dapat diikuti untuk menyelesaikan tugas atau masalah secara efisien.

Pengertian utama algoritma meliputi beberapa elemen:

1. Langkah-langkah: Algoritma terdiri dari serangkaian langkah-langkah atau instruksi yang harus diikuti untuk mencapai tujuan tertentu. Setiap langkah harus jelas, terdefinisi, dan dapat dilaksanakan.
2. Terstruktur: Algoritma harus memiliki struktur yang teratur dan sistematis. Setiap langkah harus mengikuti urutan tertentu dan harus berakhir pada hasil yang pasti.
3. Input dan Output: Algoritma mengambil input atau data dari masukan tertentu dan menghasilkan output atau solusi dari masalah setelah langkah-langkah dijalankan.
4. Efisiensi: Algoritma diharapkan memiliki kinerja yang baik, artinya ia dapat menyelesaikan masalah dengan efisien dalam waktu yang wajar dan dengan penggunaan sumber daya yang terbatas.
5. Definisi: Setiap langkah dalam algoritma harus didefinisikan dengan jelas dan dapat dimengerti oleh mesin atau manusia yang akan melaksanakannya.

Algoritma digunakan dalam berbagai bidang, termasuk pemrograman komputer, matematika, ilmu data, rekayasa, kecerdasan buatan, dan sebagainya. Membuat algoritma yang baik dan efisien sangat penting untuk mencapai hasil yang diinginkan dan mengoptimalkan proses yang melibatkan pemrosesan data dan pengambilan keputusan.

Analisis Sentimen

Analisis Sentimen adalah proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, analisis Sentimen digunakan untuk memahami komentar yang diciptakan oleh pengguna (internet) dan menjelaskan bagaimana sebuah produk maupun brand diterima oleh mereka.

Menurut Arief & Imanuel (2019) analisis sentimen juga merupakan proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif, negatif atau netral. Ide dasar dari investigasi sentimen adalah untuk mendeteksi polaritas teks dokumen atau kalimat pendek dan mengklasifikasikannya dalam premis ini. Polaritas sentimen dikategorikan sebagai "positif", "negatif" atau "tidak memihak" (netral). Proses analisis dapat

mencakup teks ulasan, forum, tweet, atau blog, dengan data *preprocessing* mencakup proses *tokenization*, *stopword*, penghapusan, *stemming*, identifikasi sentimen, dan klasifikasi sentimen.

Review Pengguna

Menurut Khammash (2008:79) dalam (Kanitra & Kusumawati, 2018). *Review* pengguna dapat dipahami sebagai salah satu media konsumen melihat *review* dari konsumen lain terhadap suatu produk, layanan perusahaan dan tentang bagaimana sebuah perusahaan produsen. Sedangkan menurut peneitian. Ulasan Online (*Online* Customer *Review*) adalah suatu ulasan yang diberikan konsumen yang berkaitan dengan informasi atas evaluasi suatu produk tentang berbagai macam aspek. Dengan informasi tersebut, konsumen bisa mendapatkan kualitas produk yang dicari dari ulasan dan pengalaman yang ditulis oleh konsumen lain yang telah membeli produk dari penjual Online sebelumnya.

Video Pendek

Menurut Rozaq & Nugrahani (2023), Video pendek biasanya memiliki durasi yang *relative* singkat, biasanya antara 15 hingga 60 detik dan memiliki arti penting dalam konten digital modern. Dalam era dimana perhatian manusia cenderung singkat, video pendek mampu menangkap dan menyampaikan pesan secara efektif dalam waktu yang terbatas. Mereka memanfaatkan daya tarik visual, narasi yang sederhana, dan pemilihan adegan yang cepat untuk menciptakan pengalaman yang menarik. Video pendek juga memungkinkan creator konten audiens yang semakin aktif di media sosial dan platform digital lainnya. Dengan format yang dioptimalkan untuk tampilan diperangkat seluler, video pendek memungkinkan pengalaman yang mudah diakses dan dikonsumsi kapan saja, dimana saja. Mereka juga mampu menghasilkan dampak yang signifikan dalam waktu yang *relative* singkat. Dengan demikian, video pendek memiliki arti penting dalam menyampaikan pesan, menghibur dan mempengaruhi audiens dunia yang semakin digital dan bergerak cepat.

Metode Algoritma K-*Nearest Neighbor*

KNN

(Cahyanti et al., 2020), *K-Nearest Neighbor (K-NN)* adalah salah satu algoritma yang sudah popular. K- NN ini termasuk ke dalam grup *instance-based, learning*. Metode K-NN merupakan teknik *lazy, learning*. Maksudnya adalah metode ini digunakan dalam klasifikasi data yang jaraknya dekat. Ada juga yang berpendapat bahwa algoritma K-NN adalah algoritma pembelajaran yang banyak digunakan dalam sistem *cyber- fisiksosial (CPSS)* untuk menganalisis dan menambang data *(main data)*

Metode Algoritma Naïve Bayes

Menurut (Kusumadewi, 2003), Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas. Pada sebuah dataset, setiap baris/dokumen diasumsikan sebagai vector dari nilai-nilai atribut dimana tiap nilai- nilai menjadi peninjauan atribut.

Naïve Bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana yang dirancang agar dapat dipergunakan dengan asumsi antar variable penjelas saling bebas (independen) dan merupakan pengklasifikasian stastistik yang dapat digunakan untuk memprediksi probabilitas keangotaan suatu class. Naïve Bayes merupakan suatu kelas keputusan, dengan menggunakan probabilitas matematika dengansyarat bahwa nilai keputusan adalah benar.

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

METODOLOGI PENELITIAN

Pada penelitian ini data yang digunakan ialah data ulasan berbahasa Indonesia yang terdapat pada Google Play Store. Penelitian ini menggunakan rapidminer sebagai alat bantu dalam proses pre-prcessing dan analisis. Analisis sistem yang diterapkan pada penelitian ini menggunakan text mining yang dikumpulkan menggunakan algoritma Metode K-NN dan Naïve bayes. Adapun proses terdiri dari beberapa tahapan analisa yaitu proses pengumpulan data dan praprosessing data.

Jenis data yang digunakan penelitian ini berasal dari data sekunder adalah sumber data penelitian yang diperoleh peneliti secara tidak langsung melalui media perantara. Dalam penelitian ini data diperoleh dengan menggunakan beberapa teknik pengumpulan data yaitu dengan mengambil data dari ulasan pengguna aplikasi video pendek (Tiktok) dengan memanfatkan *Google colab Research.* Data ulasan yang diperoleh digunakan sebagai acuan penelitian dalam menganalisa perbandingan metode *K-NN dan Naïve bayes* untuk menganalisis sentimen

Hasil analisa sistem yang berjalan untuk proses analisis digunakan adalah Kebutuhan *Software:* System operasi windows 7 ultimate 32-bit, Bahasa pemograman Rapid miner 10.1, Notepad dan *Microsoft Excel.* Kebutuhan *Hardware* yaitu Laptop Lenovo L580 core i5, Ram 8GB, *Harddisk* 256 SSD.

Dari pengolahan nilai-nilai yang ada pada kolom matriks *(True Negative (TN). False Positive (FP), False Negative (FN), dan True Postivie (TP))* maka dapat diketahui nilai *accuracy precision,* dan *recall.*

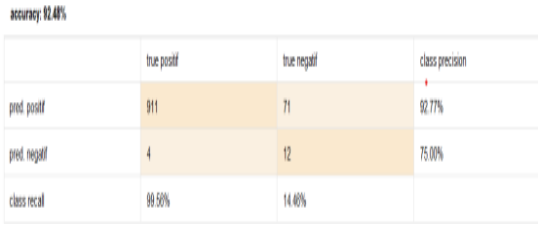
HASIL DAN PEMBAHASAN

Hasil Pengujian

Setelah proses *preprocessing*, selanjutnya akan dilakukan proses klasifikasi melalui tahap *validation:*

Pengujian Algoritma *K-Nearest Neighbor (K-NN)*

Hasil yang diperoleh dari pengujian dengan menggunakan metode *K-Nearest Neighbor*. Berikut merupakan hasil dari confusion matrix menggunakan algoritma *K- Nearest Neighbor* di tunjukan pada gambar 1.


accuracy: 92.48%

	true positif	true negatif	class precision
pred. positif	911	71	92.77%
pred. negatif	4	12	75.00%
class recall	99.56%	14.46%	

Gambar 1. Hasil Nilai *Accuracy*

1. *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x100\%$$

$$Accuracy = \frac{911 + 12}{911 + 71 + 4 + 12} \times 100\%$$

$$Accuracy = \frac{92300}{998} = 92{,}48\%$$



Gambar 2. Hasil Nilai *Precison*

2. *Precision*

$$Precison\ Positif = \frac{TP}{TP + FP} x100\%$$

$$Precision\ Positif = \frac{911}{911 + 71} \times 100\%$$

$$Precision\ Positif = \frac{91100}{982} = 92{,}77\%$$

$$Precison\ Negatif = \frac{TN}{TN + FN} x100\%$$

$$Precision\ Negatif = \frac{12}{12 + 4} \times 100\%$$

$$Precision\ Negatif = \frac{1200}{16} = 75{,}00\%$$

Gambar 3. Hasil Nilai *Recall*

3. *Recall*

$$Recall\ Positif = \frac{TP}{TP + FN} x100\%$$

$$Recall\ Positif = \frac{911}{911 + 4} x100\%$$

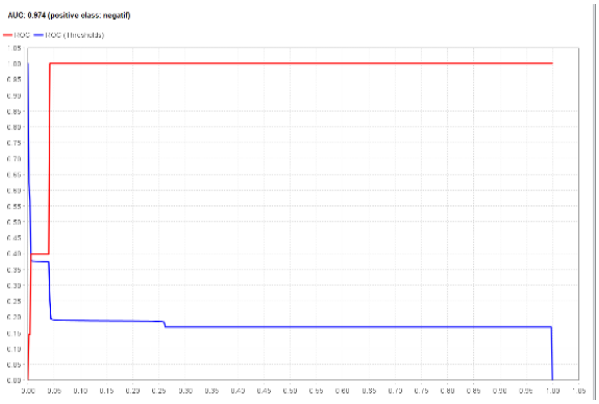
$$Recall\ Positif = \frac{91100}{915} = 99,56\%$$

$$Recall\ Negatif = \frac{TN}{TN + FP} x100\%$$

$$Recall\ Negatif = \frac{12}{12 + 71} x100\%$$

$$Recall\ Positif = \frac{1200}{83} = 14,45\%$$

Hasil akurasi pengujian K-NN untuk analisis sentiment pada Ulasan Aplikasi video pendek Tiktok memperoleh hasil sebesar 92,48%. Untuk *class precision* prediksi positif adalah 92,77% sedangkan untuk prediksi negative memperoleh hasil 75,00% dan r*ecall* yang diperoleh dari *true* positif sebesar 99,56% sedangkan untuk *true negative* sebesar 14,45%

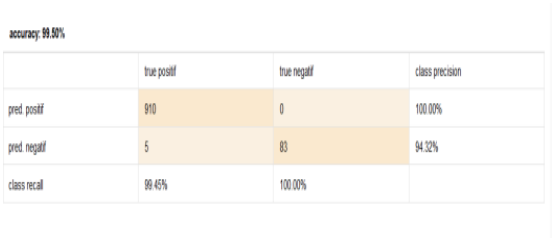


Gambar 4. Grafik Area Under Curve (AUC) Metode K-Nearest Neighbour (K-NN)

Area Under Curve (AUC) yang dihasilkan adalah 0.974 dimana pengklasifikasian keakuratan dalam penelitian ini masuk kedalam kelompok Good Classification

Pengujian Metode Naïve Bayes

Hasil yang diperoleh dari pengujian dengan menggunakan metode Naïve Bayes *Classifier*. Berikut hasil Pengujian Model menggunakan confusion matrix dengan Algoritma *Naïve Bayes Classifier* pada data



Gambar 5. Hasil Nilai Accuracy

1. *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x100\%$$

$$Accuracy = \frac{910 + 83}{910 + 83 + 5 + 0} \times 100\%$$

$$Accuracy = \frac{99300}{998} = 99{,}50\%$$

precision: 94.32% (positive class: negatif)

	true positif	true negatif	class precision
pred. positif	910	0	100.00%
pred. negatif	5	83	94.32%
class recall	99.45%	100.00%	

Gambar 6. Hasil Nilai *Precison*

2. *Precision*

$$Precison\ Positif = \frac{TP}{TP + FP} x100\%$$

$$Precision\ Positif = \frac{910}{910 + 0} \times 100\%$$

$$Precision\ Positif = \frac{9100}{910} = 100{,}00\%$$

$$Precison\ Negatif = \frac{TN}{TN + FN} x100\%$$

$$Precision\ Negatif = \frac{83}{83 + 5} \times 100\%$$

$$Precision\ Negatif = \frac{8300}{88} = 94{,}32\%$$



Gambar 7. Hasil Nilai Recall

3. *Recall*

$$Recall\ Positif = \frac{TP}{TP + FN} x100\%$$

$$Recall\ Positif = \frac{910}{910 + 5} x100\%$$

$$Recall\ Positif = \frac{910}{915} = 99{,}45\%$$

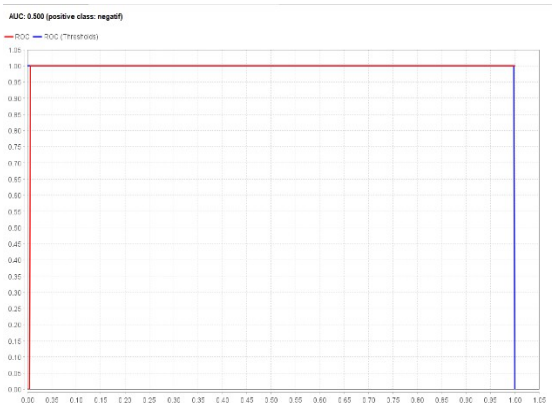
$$Recall\ Negatif = \frac{TN}{TN + FP} x100\%$$

$$Recall\ Negatif = \frac{83}{83 + 0} x100\%$$

$$Recall\ Negative = \frac{8300}{83} = 100{,}00\%$$

Hasil akurasi pengujian *Naïve Bayes*

Hasil akurasi pengujian Naïve Bayes untuk analisis sentiment pada Ulasan Aplikasi video pendek Tiktok memperoleh hasil sebesar 99,50%.untuk analisis sentiment pada Ulasan Aplikasi video pendek Tiktok memperoleh hasil sebesar 99,50%. Untuk *class precision* prediksi positif adalah 100,00%. sedangkan untuk prediksi negative memperoleh hasil 94,32% dan r*ecall* yang diperoleh dari *true* positif sebesar 99,45% sedangkan untuk *true negative* sebesar 100,00%.



Gambar 8. Grafik Area Under Curve (AUC) Metode Naïve Bayes Clasifier

Area Under Curve (AUC) yang dihasilkan adalah 0.500 dimana pengklasifikasian keakuratan dalam penelitian ini masuk kedalam kelompok *Good Classification.*

Hasil Analisa

Setelah dilakukan analisa sentiment dengan menggunakan rapid miner, maka dapat dijabarkan jumlah data awal atau data *training* menyusut seiring dilakukan berbagai macam proses, mulai dari *crawling* hingga pengklasifikasian menggunakan algotitma *K- NN* dan *Naïve Bayes* Berikut *history* data: ***Data Crawling,*** **berupa data *training* atau data mentah yang berhasil ditarik dengan *Google colab* berjumlah 1000 data ulasan aplikasi video pendek (Tiktok) di *Google play Store***

Hasil akurasi pengujian Naïve Bayes

untuk analisis sentiment pada Ulasan Aplikasi video pendek Tiktok memperoleh hasil sebesar 99,50%. Untuk *class precision* prediksi positif adalah 100,00% sedangkan untuk prediksi negative memperoleh hasil 94,32% dan r*ecall* yang diperoleh dari *true* positif sebesar 99,45% sedangkan untuk *true negative* sebesar 100,00%. Dengan uji validasi metode *K-NN* didapatkan rata-rata untuk nilai *Precision positif* 92,77% dan *Precision Negatif* 75,00% sebesar, *Recall Positif* 99,56% *dan Recall Negatif* sebesar 14,45%. Jadi data *testing* yang digunakan dalam tahap analisis sentiment ialah sebanyak 998 data serta *accuracy* yang didapatkan dengan metode K-NN ialah 92,48%.

Dengan uji validasi metode *Naïve Bayes* didapatkan rata-rata untuk nilai *Precision positif* 100,00% dan *Precision Negatif* 94,32% sebesar, *Recall Positif* 99,45% *dan Recall Negatif* sebesar 100,00%. Jadi data *testing* yang digunakan dalam tahap analisis sentiment ialah sebanyak 998 data serta *accuracy* yang didapatkan dengan metode K-NN ialah 99,50%.

KESIMPULAN

Berdasarakan dari tahapan proses yang dijelaskan pada bab sebelumnya maka dapatditarik kesimpulan dari hasil penelitian bahwa penelitian ini dengan menggunakan Google Colab sebagai Crawling data dan Rapidminer sebagai Prepocessing, Analisis sentiment, Analisa performa pada ulasan aplikasi video pendek (tiktok) di Google play store Hasil penelitian menunjukkan bahwa metode Naïve Bayes (NB) dengan representasi TF-IDF memiliki kinerja yang lebih baik dalam analisis sentimen ulasan pengguna aplikasi video pendek TikTok dibandingkan dengan metode K-Nearest Neighbor (KNN). NB menghasilkan akurasi yang lebih tinggi dan Precision yang tinggi untuk kedua kelas sentimen (positif dan negatif).

Metode K-Nearest Neighbor (KNN) menghasilkan akurasi sekitar 92,48% dalam mengklasifikasi sentimen pengguna aplikasi TikTok. Sementara itu, metode Naïve Bayes (NB) mencapai akurasi sekitar 99,50%, menunjukkan performa yang jauh lebih baik dalam memprediksi sentiment ulasan dikarenakan faktor nilai True negatif pada Naïve Bayes lebih kecil dibandingkan dengan K-NN semakin kecil true negatif semakin besar akurasi yang di dapatkan.

Tingkat akurasi keseluruhan dalam mengklasifikasi sentimen pengguna aplikasi video pendek TikTok adalah sekitar 99,50%, yang mengindikasikan kemampuan tinggi dalam membedakan antara ulasan positif dan negatif. Hasil analisis juga menunjukkan bahwa metode Naïve Bayes (NB) memiliki Precision positif sebesar 100,00%, yang berarti kemampuan yang sangat baik dalam mengklasifikasikan ulasan sebagai sentimen positif.

DAFTAR PUSTAKA

Ahmadi, M. I., Apriani, F., Kurniasari, M., Handayani, S., & Gustian, D. (2020). Sentiment Analysis Online Shop on the Play Store Using Method Support Vector Machine (Svm). Seminar Nasional Informatika (SEMNASIF), 1(1), 196–203.

Arief, R., & Imanuel, K. (2019). Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter dengan Metode Lexicon Based. Jurnal Ilmiah Matrik, 21(3), 242–250.

Basah, S. (1994). Hukum Tata Negara Perbandingan. Bina Aksara.

Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis Performa Metode KNN pada Dataset Pasien Pengidap Kanker Payudara. Indonesian Journal of Data and Science, 1(2), 39–43.

Candra, R. M., & Rozana, A. N. (2020). Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor. *IT Journal Research and Development*, *5*(1), 45–52.

Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, *8*(6), 219–225.

Elmasri, R., & Navathe, S. B. (2004). *Database Systems: models, languages, design, and application programming.* Pearson Education India.

Hoffer, J. A. (2005). *Modern Database Management*. Pearson.

Hoffer, J. A., George, J. F., & Valacich, J. S. (2004). *Modern System Analysis and Design*. Pearson.
Kambey, G. E. I., Sengkey, R., & Jacobus, A. (2020). Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia. *Jurnal Teknik Informatika*, *15*(2), 75–82.

Kanitra, A. R., & Kusumawati, A. (2018). Pengaruh Country of Origin dan Online Consumer Review Terhadap Trust dan Keputusan Pembelian (Survei pada Mahasiswa S1 Fakultas Ilmu Administrasi Universitas Brawijaya Angkatan 2015/2016 dan 2016/2017 Tahun Akademik 2017/2018 Pembeli Produk Oppo Smartphon. Jurnal Administrasi Bisnis, 61(1), 64–73.

Kusumadewi, S. (2003). Artificial Intelligence (Teknik dan Aplikasinya). Graha Ilmu.
Lestari, S., & Saepudin, S. (2021). Support Vector Machine: Analisis Sentimen Aplikasi Saham di Google Play Store. JUSIFO (Jurnal Sistem Informasi), 7(2), 81–90.

Marhiyanto, B. (2000). Kamus Lengkap Bahasa Indonesia. Media Center.
Mo, Z., Li, Y.-F., & Fan, P. (2015). Effect of Online Reviews on Consumer Purchase Behavior. Journal of Service Science and Management, 8(03), 419. ttps://doi.org/10.4236/jssm.201 5.83043

Musfiroh, D., Khaira, U., Eko, P., Utomo, P., & Suratno, T. (2021). Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 1(1), 24–33.

Nugroho, A. (2013). Klasifikasi Naïve Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil. BIMIPA, 23(3), 297–308.

Nurwidyantoro, A., & Winarko, E. (2012). Parallelization of maximum entropy POS tagging for Bahasa Indonesia with MapReduce. International Journal of Computer Science Issues (IJCSI), 9(4). https://doi.org/10.48550/arXiv.12 08.3047

Perkovis, L. (2012). *Introduction to Computing Using Python: an application development focus*. Wiley.

Rasenda, R., Lubis, H., & Ridwan, R. (2020). Implementasi K-NN dalam Analisa Sentimen Riba pada Bunga Bank Berdasarkan Data Twitter. *Jurnal Media Informatika Budidarma*, *4*(2), 369–376.

Rashel, F., & Manurung, R. (2014). Pemuisi: A Constraint Satisfaction-Based Generator Of Topical Indonesian Poetry. *Proceedings of 5th International Conference on Computational Creativity (ICCC)*, 82–90.

Rozaq, M., & Nugrahani, R. U. (2023). Penggunaan Platform Video Pendek Sebagai Strategi Komunikasi Pemasaran Digital untuk UMKM. *Jurnal Komunikasi Nusantara*, *5*(1), 21–30.

Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, *5*(1), 157–163.

Sholihin, A., Haviluddin, H., Puspitasari, N., Wati, M., & Islamiyah, I. (2019). Analisis Penyakit Difteri Berbasis Twitter Menggunakan Algoritma Naïve Bayes. *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, *1*(1), 7–15.

Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Alfabeta.

Syafnidawaty. (2020). *Perbedaan Metodologi Penelitian dan Metode Penelitian*. Universitas Raharja. Verawardina, U., Asnur, L., Lubis, A. L., Hendriyani, Y., Ramadhani, D., Dewi, I. P., Darni, R., Betri, T. J., Susanti, W., & Sriwahyuni, T. 2020). Reviewing Online Learning Facing the Covid-19 Outbreak. *Talent Development and Excellence*, *12*(SpecialIssue3), 385–392.

Wati, F. F. (2020). *Analisis Sentimen Review Aplikasi Tik-Tok dengan Algoritma K-Nearest Neighbor, Naive Bayes dan Support Vector Machine*. Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

Wibowo, J. T. P. (2021). *Apa Itu Google Colaboraty*. Warta Ekonomi.

Widianto, M. H. (2019). *Algoritma Naive Bayes*. binus.ac.id. https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes