

**ANALISIS SENTIMEN MEDIA SOSIAL TERHADAP PENGGUNAAN
INSTAGRAM OLEH ANAK USIA DINI PENGEMBANGAN SISTEM
INFORMASI EDUKATIF BERBASIS *NAÏVE BAYES***

SKRIPSI

Diajukan sebagai salah satu syarat kelulusan program Strata-1 (S-1)

Program Studi: Sistem Informasi

Tahun Akademik 2024 / 2025



**STIKOM
CIPTA KARYA INFORMATIKA**

Disusun Oleh:

Nama : Salsabila Putri W

NIM : 21120110025

**SEKOLAH TINGGI ILMU KOMPUTER CIPTA KARYA INFORMATIKA
(STIKOM CKI) JAKARTA**

2025



PENGESAHAN UJIAN

SKRIPSI ini diujikan pada tanggal 16 bulan Agustus tahun 2025 , dan dinyatakan : ~~LULUS~~/TIDAK LULUS

Nama : Salsabila Putri W

Nim : 21120110024

Program Studi/Program : Sistem Informasi/S-1

Judul SKRIPSI : Implementasi model Naive Bayes multikategori untuk analisis sentimen produk Wardah di E-commerce Shoppe

NAMA PENGUJI :

TANDA TANGAN

1. Yuma Akbar, S.Kom., M.Kom
NIDN : 0319038304

2. Prof. Dr. H.Supriyatin, SY., MM
NIDN : 8815080018

Mengetahui, Ketua Sidang SKRIPSI

(Dr. Mesra Betty Yel, MM, DBA M. Kom)

NIDN : 0416057202



**SEKOLAH TINGGI ILMU KOMPUTER CIPTA KARYA
INFORMATIKA**

SURAT KETERANGAN KEASLIAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, SKRIPSI ini merupakan karya saya sendiri (ASLI), dan isi dalam SKRIPSI ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dibuktikan dengan cek plagiasi Turnitin dengan plagiasi kurang dari 30% yang disetujui oleh Kepala LPPM STIKOM CKI.

Berikut menerangkan kembali bahwa saya:

NIM : 21120110025

Nama : Salsabila Putri W

Judul SKRIPSI : Analisis Sentimen Media Sosial Terhadap Penggunaan Instagram Oleh Anak Usia Dini Pengembangan Sistem Informasi Edukatif Berbasis Naïve Bayes

Adalah benar karya Tulis saya dan segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Jakarta, 13 Agustus 2025

Mahasiswa,



Salsabila Putri W
NIM:21120110025

ABSTRAK

Penelitian ini menggunakan metode Naïve Bayes untuk menganalisis sentimen media sosial terkait penggunaan Instagram oleh anak usia dini. Tujuan utama penelitian adalah memahami persepsi publik mengenai dampak positif dan negatif dari penggunaan Instagram pada anak usia dini, khususnya pada aspek perkembangan sosial, psikologis, dan perilaku digital mereka. Analisis sentimen dilakukan dengan memanfaatkan data dari berbagai platform media sosial yang kemudian diklasifikasikan menjadi opini positif, negatif, dan netral. Hasil klasifikasi ini menjadi dasar dalam pengembangan sistem informasi edukatif terintegrasi yang dirancang untuk memberikan panduan bagi orangtua, pendidik, serta anak-anak agar mampu menggunakan Instagram secara aman, sehat, dan bertanggung jawab. Sistem ini juga menekankan pada pentingnya edukasi konten sesuai usia, pengaturan privasi, serta strategi untuk meminimalkan risiko paparan konten yang tidak sesuai maupun dampak penggunaan berlebihan. Penelitian ini diharapkan dapat mendukung terciptanya lingkungan digital yang lebih positif, aman, dan bermanfaat bagi anak usia dini sekaligus menjadi acuan dalam merumuskan kebijakan yang efektif di era media sosial.

Kata Kunci : Metode *Naïve Bayes*, edukatif, Media sosial *Instagram*, anak usia dini.

ABSTRACT

This study employs the Naïve Bayes method to analyze social media sentiment regarding the use of Instagram by early childhood users. The primary objective of this research is to understand public perceptions of the positive and negative impacts of Instagram usage among young children, particularly in relation to their social, psychological, and digital behavioral development. Sentiment analysis is carried out using data from various social media platforms, which are then classified into positive, negative, and neutral opinions. The classification results form the basis for developing an integrated educational information system designed to provide guidance for parents, educators, and children in using Instagram safely, healthily, and responsibly. The system also emphasizes the importance of age-appropriate content education, privacy settings, and strategies to minimize the risks of exposure to inappropriate content and the negative effects of excessive usage. This research is expected to support the creation of a more positive, safe, and beneficial digital environment for early childhood users while also serving as a reference in formulating effective policies in the social media era.

Keywords: *Naïve Bayes Method, educational, Instagram social media, early childhood.*

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan hidayah-Nya karena berkat Rahmat dan kasih nya, saya dapat menyelesaikan Tugas Akhir Skripsi dengan judul **“Analisis Sentimen Media Sosial terhadap Penggunaan Instagram oleh Anak Usia Dini: Pengembangan Sistem Informasi Edukatif Berbasis *Naïve Bayes*”**, dapat disusun dengan sebagaimana mestinya. Skripsi ini dapat diselesaikan dengan bantuan dan Kerjasama oleh berbagai pihak, pada kesempatan ini saya ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Muhammad Farel Ardhan, S.T., MBA selaku ketua Yayasan Cipta Karya Intelektual.
2. Ibu Dr. Mesra Betty Yel, MM, MBA., M.Kom selaku Ketua Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI).
3. Bapak Yuma Akbar, S.kom., M.Kom selaku Wakil Ketua I Bidang Akademik dan Kemahasiswaan Sekolah Tinggi Komputer Cipta Karya Informatika (STIKOM CKI).
4. Bapak Muchammad Zaeny, SE., M.Pd selaku Wakil Ketua II Bidang SDM, Keuangan dan Sarana Prasana Sekolah Tinggi Komputer Cipta Karya Informatika (STIKOM CKI).
5. Bapak Muhamad Harun, M.Pd selaku Kepala LPM Sekolah Tinggi Komputer Cipta Karya Informatika (STIKOM CKI).
6. Bapak Dr. Suhendi, S.T.,S.Kom., M.Msi selaku Ketua LPPM Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
7. Bapak Veri Arinal, S.Kom, M.Kom selaku Ketua Program Studi Sistem Informasi Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
8. Bapak Dadang Iskandar Mulyana, S.Kom, M.Kom selaku ketua Program Studi Teknik Infromatika Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
9. Ibu Sri Lestari, S.pd., MM selaku ketua Program Studi Bisnis Digital Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
10. Bapak Veri Arinal, S.Kom, M.Kom selaku Dosen Pembimbing yang telah banyak memberikan masukan, semangat, serta dorongan selama penyusunan skripsi.

11. Bapak Ir. Sugiono, M.Kom selaku dosen Pembimbing Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
12. Terimakasih kepada penguji 1 Bapak Yuma Akbar, S.Kom., M.Kom Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
13. Terimakasih kepada penguji 2 Bapak Prof. Dr. H.Supriyatin, SY., MM Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
14. Management dan Seluruh Dosen Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika.
15. Terimakasih kepada teman-teman sejawat
16. Ucapan terima kasih juga disampaikan kepada pihak Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika atas dukungan teknis dan fasilitas yang diberikan dalam pengembangan media pembelajaran ini.
17. Orang Tua yang selalu mendoakan, memberikan dukungan, semangat, dan mendoakan saya dalam proses Skripsi ini.
18. Ucapan terimakasih kepada diri saya sendiri yang telah berusaha semaksimal mungkin, terimakasih karena sudah bertahan, terimakasih karena tidak menyerah dan terimakasih karena selalu percaya pada diri sendiri.

Jakarta, 20 Mei 2025

Penulis Skripsi



Salsabila Putri W

DAFTAR TABEL

tabel 2.1	Review PICOC	7
tabel 2.2	Review Survei Protocol.....	8
tabel 2.3	Kajian Pustaka Utama	10
tabel 2.4	Matrix Perbandingan RP,RQ, dan RO	35
tabel 2.5	Flowchart.....	42
tabel 4.1	Spesifikasi Software	50
tabel 4.2	Spesifikasi Hardware.....	51

DAFTAR GAMBAR

gambar 2.1	Studies Selection Strategy	9
gambar 3.1	Tahapan Penerapan Metodologi	46
gambar 3.2	Kerangka Penelitian.....	47
gambar 4.1	jumlah data yang di dapat.....	52
gambar 4.2	Model <i>Proses Cleansing</i>	53
gambar 4.3	Hasil <i>Cleansing Data</i>	53
gambar 4.4	data yang sudah dilabeli	54
gambar 4.5	Model <i>Tokenizing Data</i>	55
gambar 4.6	Hasil <i>Tokenizing Naïve bayes</i>	55
gambar 4.7	Deskripsi Akurasi <i>Naïve bayes</i>	55
gambar 4.8	Deskripsi <i>Simple Distribution</i>	56
gambar 4.9	<i>Plot View Naïve bayes</i>	57
gambar 4.10	<i>Table View</i>	57
gambar 4.11	<i>Result history</i>	58
gambar 4.12	<i>Distribution table</i>	58

DAFTAR ISI

ABSTRAKSI.....	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR.....	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR ISI.....	x
DAFTAR PENELITIAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah (<i>Research Problem / RP</i>).....	2
1.3 Rumusan Masalah (<i>Research Question / RQ</i>).....	2
1.4 Tujuan Penelitian (<i>Research Objective / RO</i>)	3
1.5 Pendekatan Pemecahan Masalah.....	3
1.6 State Of The Art (SOTA)	4
1.7 Kontribusi Penelitian/Novelty	5
BAB II TINJAUAN PUSTAKA	6
2.1 Systematic Literature Review (SLR)	6
2.1.1 Survei Metodologi	7
2.1.2 . <i>Review Survey Protocol</i>	8
2.1.3 Kajian Pustaka Utama	9
2.1.4 Matrik Perbandingan RP,RQ, dan RO	34
2.2 Definisi dan Pengertian	37
2.2.1 Anak usia dini menggunakan media sosial	37
2.2.2 Analisis Sentimen.....	37

2.2.3 Instagram	38
2.2.4 <i>Naïve Bayes</i>	38
2.2.5 Data	39
2.2.6 Rapid Minner.....	40
2.2.7 Media Sosial	41
2.2.8 Pengembangan Sistem.....	41
2.2.9 <i>Flowchart</i>	42
BAB III METODE PENELITIAN	45
3.1 Data Penelitian.....	45
3.2 Penerapan Metode	45
3.3 Rancangan Pengujian	48
3.3.1 Business Understanding (Pemahaman Bisnis)	48
3.3.2 Data Understanding (Pemahaman Data)	48
3.3.3 Data Preparation (Persiapan Data)	48
3.3.4 Modeling (Pemodelan)	49
3.3.5 Evaluation (Evaluasi)	49
3.3.6 Deployment	49
BAB IV HASIL YANG DIHARAPKAN.....	50
4.1 Alat Penelitian	50
4.1.1 Perangkat lunak (<i>software</i>).....	50
4.1.2 Perangkat Keras (<i>hardware</i>)	50
4.2 Implementasi Dan Pengujian.....	51
4.3 Results	55
BAB V.....	60
PENUTUP.....	60
5.1 Kesimpulan.....	60

5.2 Saran.....	61
Daftar Pustaka	62

DAFTAR PENELITIAN

No.	Bln-Th Terbit	Judul Penelitian Anda	Link URL Publikasi	Jurnal Nasional	Jurnal Internasional
1.	30 Juni 2025	Implementasi Sistem Informasi Perkembangan Balita Berbasis Website Di Posyandu Melati RW 2 Kelurahan Munjul	https://ejournal.nusantaraglobal.ac.id/index.php/devote/article/view/4113	Terindex Sinta (4)	-

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah membawa perubahan signifikan dalam interaksi sosial manusia di era modern. Media sosial, sebagai salah satu bentuk teknologi komunikasi, telah mengalami perkembangan pesat dan menjadi bagian tak terpisahkan dari kehidupan sehari-hari masyarakat. Instagram, sebagai salah satu platform media sosial yang menekankan pada konten visual, telah mencapai popularitas luar biasa dengan lebih dari 2 miliar pengguna aktif bulanan di seluruh dunia per tahun 2024.[1]

Penggunaan Instagram oleh anak usia dini menimbulkan berbagai kekhawatiran dan pendapat di masyarakat. Beberapa pihak menekankan potensi dampak negatif seperti paparan konten tidak sesuai, risiko *cyberbullying*, dan gangguan perkembangan sosial. Di sisi lain, terdapat pandangan bahwa dengan pengawasan yang tepat.[2]

Analisis sentimen, merupakan cabang ilmu komputer yang bertujuan untuk mengidentifikasi, mengekstrak, dan mengkuantifikasi pendapat serta sikap seseorang terhadap entitas, peristiwa, atau topik tertentu. Dalam konteks media sosial, analisis sentimen dapat digunakan untuk memetakan opini publik terhadap penggunaan Instagram oleh anak usia dini. Hasil dari analisis ini dapat memberikan wawasan berharga bagi pengembangan kebijakan, program edukasi, serta sistem informasi yang dapat membantu orang tua dan pendidik dalam mengarahkan penggunaan media sosial secara sehat bagi anak-anak.[3]

Metode klasifikasi *Naïve Bayes* dipilih dalam penelitian ini karena keunggulannya dalam analisis sentimen teks. Algoritma ini terkenal dengan efisiensi komputasinya, kemampuan bekerja dengan dataset yang relatif kecil, serta keefektifannya dalam klasifikasi teks berbahasa Indonesia. Melalui pendekatan probabilistik, metode *Naïve Bayes* dapat mengklasifikasikan opini publik ke dalam sentimen positif, negatif, atau netral berdasarkan fitur linguistik yang terdapat dalam dataset. [4]

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan sistem informasi edukatif berbasis analisis sentimen menggunakan metode *Naïve Bayes*. Sistem ini diharapkan dapat memberikan informasi komprehensif mengenai pandangan masyarakat terhadap penggunaan Instagram oleh anak usia dini, sekaligus menjadi platform edukasi bagi orang tua dan pendidik dalam mendampingi anak menggunakan media sosial secara sehat dan bertanggung jawab.

1.2 Identifikasi Masalah (*Research Problem / RP*)

Berdasarkan uraian latar belakang diatas, maka peneliti mengidentifikasi masalah yaitu :

- a) **Peningkatan Penggunaan Media Sosial oleh Anak Usia Dini:** Terdapat peningkatan signifikan dalam penggunaan media sosial, khususnya Instagram, oleh anak-anak di bawah usia 13 tahun, yang melanggar batasan usia yang telah ditetapkan oleh platform tersebut.
- b) **Dampak Negatif dari Penggunaan Instagram:** Penggunaan Instagram oleh anak-anak dapat menimbulkan berbagai dampak negatif, termasuk paparan konten yang tidak sesuai, risiko cyberbullying, serta gangguan dalam perkembangan sosial mereka.
- c) **Kurangnya Pemahaman dan Pengawasan dari Orang Tua dan Pendidik:** Masyarakat, termasuk orang tua dan pendidik, mungkin tidak memiliki pemahaman yang cukup mengenai bagaimana mengawasi dan mendampingi anak-anak dalam menggunakan media sosial secara sehat dan bertanggung jawab.

1.3 Rumusan Masalah (*Research Question / RQ*)

Berdasarkan identifikasi masalah di atas, rumusan masalah dalam penelitian ini adalah :

- a. Faktor-faktor apa yang mempengaruhi penggunaan Instagram oleh anak usia dini di Indonesia?
- b. Bagaimana pandangan dan opini masyarakat tentang dampak penggunaan Instagram pada anak di bawah 13 tahun?

- c. Jenis konten apa yang diakses anak-anak di Instagram, dan bagaimana dampaknya pada perkembangan sosial dan psikologis mereka?

1.4 Tujuan Penelitian (*Research Objective / RO*)

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a) Menganalisis Penggunaan Instagram oleh Anak Usia Dini: Untuk memahami faktor-faktor yang mempengaruhi peningkatan penggunaan Instagram di kalangan anak-anak di bawah usia 13 tahun.
- b) Mengidentifikasi Opini Publik: Untuk memetakan dan mengidentifikasi pandangan serta opini masyarakat mengenai dampak penggunaan Instagram oleh anak usia dini, baik yang positif maupun negatif.
- c) Mengedukasi Orang Tua dan Pendidik: Untuk menyediakan platform edukatif yang membantu orang tua dan pendidik memahami penggunaan media sosial oleh anak-anak dan memberikan panduan dalam pengawasan dan pendampingan mereka.

1.5 Pendekatan Pemecahan Masalah

Penelitian ini adalah pengembangan sistem informasi edukatif berbasis analisis sentimen. Penelitian ini menggunakan metode analisis sentimen untuk mengidentifikasi opini publik terkait penggunaan Instagram oleh anak usia dini. Analisis sentimen ini akan menggunakan algoritma *Naïve Bayes*[2], yang dipilih karena efisiensi komputasinya dan keefektifannya dalam mengolah teks berbahasa Indonesia. Hasil analisis sentimen kemudian akan diintegrasikan ke dalam sebuah sistem informasi edukatif yang bertujuan memberikan informasi komprehensif dan edukasi kepada orang tua dan pendidik tentang penggunaan media sosial yang sehat dan bertanggung jawab. Dengan kata lain, penelitian ini menggunakan pendekatan kuantitatif (analisis sentimen) untuk menghasilkan data yang kemudian diinterpretasikan dan disajikan dalam bentuk sistem informasi edukatif yang bersifat kualitatif.[5]

1.6 State Of The Art (SOTA)

Berikut adalah beberapa penelitian terdahulu yang menjadi acuan dalam pengembangan sistem ini :

- a. Research on the use of data mining for business strategy in the culinary sector continues to grow, particularly through the combination of RFM analysis and the Apriori algorithm. RFM is used for customer segmentation based on transaction behavior, while Apriori identifies product combinations that are frequently purchased together. Previous studies have demonstrated the effectiveness of these methods, such as Merliani et al. (2022) in menu recommendation, and Susilowati et al. (2022) and Widiyanto & Witanti (2021), who combined them with clustering techniques. In this study, the integrated RFM and Apriori approach was applied at Bonjour Café and Resto using transaction data from December 2024. A strong association was found between "Grilled Tilapia," "Vit 550 ml," and "French Fries" (support 28%, confidence 70%), along with the segmentation of five customer types. The main contribution of this study lies in the combined application of RFM–Apriori with a monthly focus in the culinary industry, which is rarely explored in existing literature, and the use of the CRISP-DM approach to ensure systematic and valid analysis.[6]
- b. Penelitian ini menganalisis sentimen masyarakat di Twitter terkait penutupan TikTokShop oleh pemerintah. Menggunakan algoritma Naïve Bayes dan data berlabel dari *VADER lexicon*, studi ini menunjukkan bahwa mayoritas netizen memberikan respon positif terhadap kebijakan tersebut. Model yang dibangun menunjukkan akurasi 85%, dengan precision 99,33% dan recall 77,20%. Studi ini memberikan kontribusi pada analisis opini publik digital serta relevan untuk pengambilan kebijakan berbasis data sosial.[7]

- c. Penelitian ini menganalisis opini publik terhadap kaum LGBT di media sosial Twitter menggunakan algoritma Naïve Bayes. Dari 1.579 tweet yang dikumpulkan, sebanyak 1502 diklasifikasikan sebagai negatif dan 77 sebagai positif. Model mencapai akurasi 95,24% dan 87,58% dengan validasi silang. Hasil ini menunjukkan dominasi opini negatif di Twitter terhadap LGBT, sekaligus membuktikan efektivitas metode Naïve Bayes dalam klasifikasi opini sosial.[8]

1.7 Kontribusi Penelitian/Novelty

Penelitian ini memberikan kontribusi orisinal dengan mengembangkan sistem informasi edukatif berbasis analisis sentimen menggunakan metode *Naïve Bayes* untuk memetakan opini publik mengenai penggunaan Instagram oleh anak usia dini. Keunikannya terletak pada penerapan metode ini dalam konteks spesifik ini, yang belum banyak diteliti sebelumnya.[9]

Meskipun analisis sentimen telah digunakan dalam berbagai konteks, aplikasinya pada pemetaan opini publik tentang penggunaan media sosial oleh anak-anak, khususnya Instagram, dan penyediaan sistem informasi edukatif berbasis hasil analisis tersebut, merupakan hal yang relatif baru. Penelitian sebelumnya mungkin telah menganalisis sentimen terkait media sosial atau penggunaan internet oleh anak-anak, tetapi jarang yang menggabungkan analisis sentimen dengan pengembangan sistem informasi edukatif yang ditujukan langsung kepada orang tua dan pendidik.[10]

Sistem informasi yang dihasilkan diharapkan dapat memberikan informasi komprehensif dan mudah dipahami mengenai persepsi masyarakat terhadap dampak positif dan negatif penggunaan Instagram oleh anak-anak. Informasi ini dapat membantu orang tua dan pendidik dalam membuat keputusan yang lebih tepat dan terinformasi terkait pengawasan dan bimbingan penggunaan media sosial oleh anak-anak di bawah umur. Dengan demikian, penelitian ini berkontribusi pada peningkatan literasi digital dan pengasuhan anak di era digital.[11]

BAB II

TINJAUAN PUSTAKA

2.1 Systematic Literature Review (SLR)

Systematic literature review atau sering di singkat SLR atau dalam bahasa indonesia disebut tinjauan pustaka sistematis adalah metode literature review yang mengidentifikasi, menilai, dan menginterpretasi seluruh temuan-temuan pada suatu topik penelitian, untuk menjawab pertanyaan penelitian (*research question*) yang telah ditetapkan sebelumnya . Secara umum tahapan untuk melakukan SLR terdiri ada 3 bagian besar, yaitu :

a. *Planning*

Research Question (RQ) adalah bagian awal dan dasar berjalannya SLR. RQ digunakan untuk menuntun proses pencarian dan ekstraksi literatur. Analisis dan sintesis data, sebagai hasil dari SLR, adalah jawaban dari RQ yang kita tentukan di depan. RQ yang baik adalah yang bermanfaat, terukur, arahnya ke pemahaman terhadap *state-of- the-art research* dari suatu topik penelitian.

b. *Conducting*

Tahapan *conducting* adalah tahapan yang berisi pelaksanaan dari SLR, dimana seharusnya sesuai dengan Protokol SLR yang telah kita tentukan. Dimulai dari penentuan *keyword* pencarian literatur (*search string*) yang basisnya adalah dari PICOC yang telah kita desain di depan. Pemahaman terhadap sinonim dan alternatif pengganti kata akan menentukan akurasi pencarian literatur kita. Kemudian langkah berikutnya adalah penentuan sumber (*digital library*) dari pencarian literatur.

c. *Reporting*

Reporting adalah tahapan penulisan hasil SLR dalam bentuk tulisan, baik untuk dipublikasikan dalam bentuk paper ke jurnal ilmiah atau untuk menyusun Bab 2 tentang *Literature Review* dari skripsi/tesis/disertasi.

2.1.1 Survei Metodologi

Menurut para ahli yaitu Kerlinger, menyatakan bahwa penelitian survei Metodologi adalah penelitian yang dilakukan pada populasi besar maupun kecil, tetapi data yang dipelajari adalah data dari sampel yang diambil dari populasi tersebut, untuk menemukan kejadian-kejadian relatif, distribusi, dan hubungan-hubungan antar variabel sosiologis maupun psikologis. [12] Dalam Perumusan RQ tentang efektivitas dalam penelitian harus fokus pada 5 elemen yang dikenal sebagai PICOC :

- a. *Population* (P) : kelompok sasaran untuk penyelidikan (misalnya orang, perangkat lunak, dll.)
- b. *Intervention* (I) : menentukan aspek penyelidikan atau masalah yang menarik bagi peneliti.
- c. *Comparison* (C) : Aspek dari investigasi dimana Intervention akan dibandingkan.
- d. *Outcomes* (O) : Efek dan hasil dari Intervention.
- e. *Context* (C) : latar atau lingkungan penyelidikan.

tabel 2.1 Review PICOC

Judul Tema Survei : ANALISIS SENTIMEN MEDIA SOSIAL TERHADAP PENGGUNAAN INSTAGRAM OLEH ANAK USIA DINI PENGEMBANGAN SISTEM INFORMASI EDUKATIF BERBASIS NAÏVE BAYES	
<i>Population</i>	Analisis Sentimen Menggunakan sistem edukatif berbasis <i>Naïve Bayes</i>

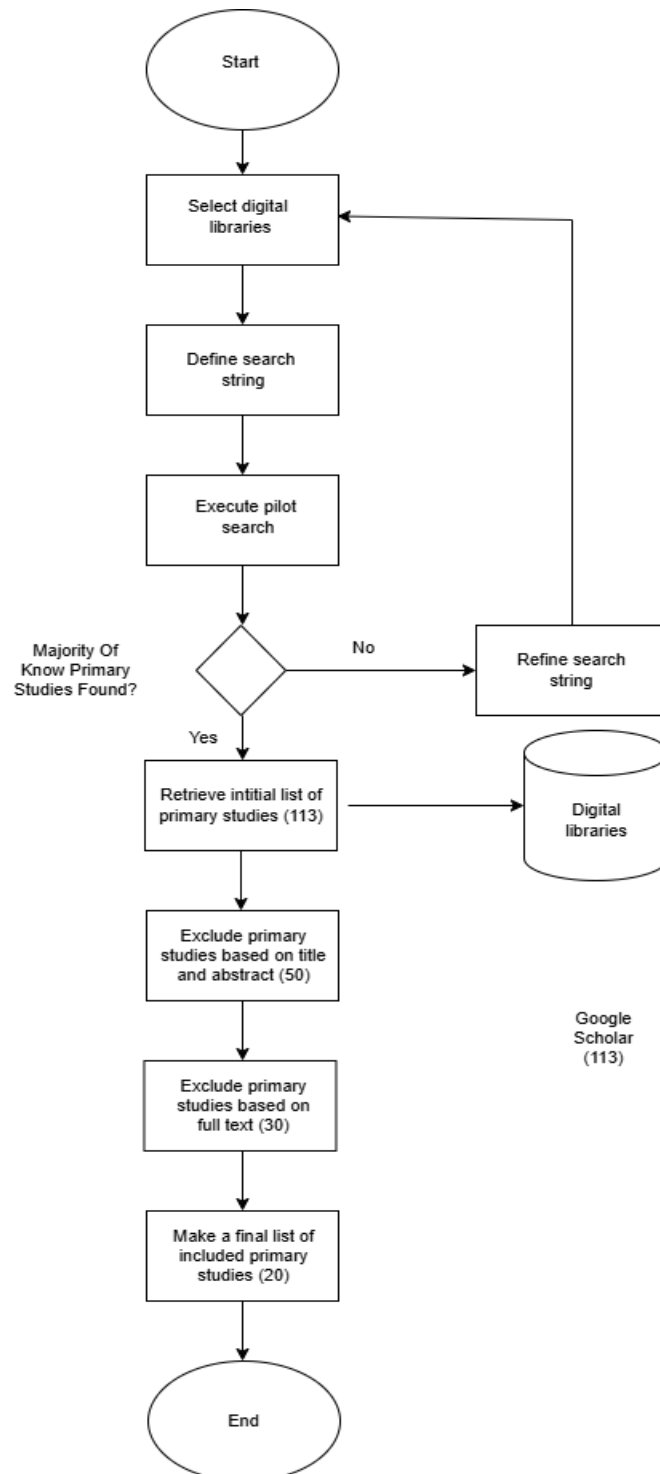
<i>Intervention</i>	Bagaimana cara peneliti dapat menentukan sentimen negatif dan positif dari masyarakat melalui sosial media Instagram
<i>Comparison</i>	<i>N/A</i>
<i>Outcomes</i>	Menerapkan konsep data mining menggunakan Algoritma K Nearest Neighbor untuk menentukan dan mengetahui sentimen masyarakat tentang Naturalisasi Tim Nasional Indonesia U-23 Di Era Shin Tae-Yong
<i>Context</i>	<i>Public</i>

2.1.2 . *Review Survey Protocol*

Proses Dimana anda melakukan metodologi survei dalam pengumpulan jurnal

tabel 2.2 Review Survei Protocol

<i>REVIEW SURVEI PROTOCOL</i>	
<i>Publication Year</i>	2022 s/d 2025
<i>Publication Type</i>	google scholar
<i>Search String</i>	(analisis sentimen, <i>Naïve Bayes</i> , instagram)
<i>Final Selected</i>	20



gambar 2.1 Studies Selection Strategy

2.1.3 Kajian Pustaka Utama

Kajian pustaka merupakan data pendukung yang wajib dijadikan

sebagai acuan yang didapat dari penelitian terdahulu yang berhubungan dengan penelitian ini. berikut ini merupakan beberapa hasil penelitian yang berhubungan dengan penelitian ini:

tabel 2.3 Kajian Pustaka Utama

KAJIAN PUSTAKA UTAMA			
No	Judul Paper	Nama Penulis	Nama Jurnal & No. ISSN / ISBN / DOI
1	ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP ROKOK ELEKTRIK (VAPE) DI INDONESIA MENGGUNAKAN METODE <i>NAÏVE BAYES</i>	Dio Rizki Aditya, Endang Supriyati, Tri Listyorini	E-ISSN : 2540 - 8984 JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Volume 07, Nomor 01, Maret 2022 : 43 – 50
	<p>Masalah Yang Ditemukan:</p> <p>Pengumpulan data dilakukan dengan memperoleh data sentimen yang relevan dan dalam jumlah yang memadai dari Twitter guna mendukung analisis yang akurat. Setelah data terkumpul, tahap selanjutnya adalah pemrosesan data (preprocessing), di mana data mentah berupa tweet dibersihkan dari noise seperti tanda baca, URL, emotikon, dan kata-kata yang tidak relevan, kemudian diproses agar menjadi format yang terstruktur dan sesuai untuk digunakan dalam algoritma Naïve Bayes.</p>		

	<p>Solusi Yang Dilakukan:</p> <p>Pengumpulan data (scraping) dilakukan dengan memanfaatkan Twitter API dan library Tweepy menggunakan bahasa pemrograman Python untuk mengambil tweet yang relevan dengan kata kunci "vape". Setelah model Naïve Bayes diterapkan, evaluasi hasil dilakukan menggunakan confusion matrix dengan mengukur performa model berdasarkan metrik akurasi, precision, recall, dan F1-score.</p>		
2	<p>SENTIMEN ANALISIS BELAJAR ONLINE DI TWITTER MENGGUNAKAN AN <i>NAÏVE</i> <i>BAYES</i></p>	<p>Oktavia Putri Zusrotun, Alif Catur Murti, Rina Fiati</p>	<p>ISSN 2089-8673 ISSN 2548-4265 Volume 11, Nomor 3, Desember 2022</p>
	<p>Masalah Yang Ditemukan:</p> <p>Infrastruktur dan biaya internet menjadi salah satu kendala utama dalam pembelajaran daring, karena membutuhkan akses internet yang memadai dan biaya yang tidak sedikit, sehingga menyulitkan sebagian peserta didik. Selain itu, komunikasi secara daring juga sering mengalami hambatan atau keterlambatan, yang dapat mengurangi efektivitas proses pembelajaran.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini menggunakan media sosial Twitter sebagai sarana untuk mengetahui opini masyarakat mengenai pembelajaran daring, karena platform ini memungkinkan siswa menyampaikan keluhan dan pendapat mereka secara efektif dan efisien. Untuk menganalisis sentimen dari tweet yang telah dikumpulkan, digunakan algoritma Naïve Bayes yang mampu mengklasifikasikan opini menjadi kategori positif, negatif, atau netral secara akurat.</p>		

3	Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode	Widya Wahyuni	Jurnal Informatika Ekonomi Bisnis http://www.infeb.org 2022 Vol. 4 No. 4 Hal: 148-153 e-ISSN: 2714-8491
<p>Masalah Yang Ditemukan:</p> <p>Data mentah dari Twitter berupa tweet bersifat tidak terstruktur sehingga perlu diproses terlebih dahulu sebelum digunakan untuk analisis sentimen. Selain itu, data tersebut juga mengandung noise, yaitu informasi yang tidak relevan seperti mention, hashtag, URL, dan tanda baca, yang harus dibersihkan agar tidak mengganggu akurasi dalam proses analisis.</p>			
<p>Solusi Yang Dilakukan:</p> <p>Tahapan preprocessing data dilakukan untuk membersihkan dan menata data tweet, mencakup proses case folding, cleansing, stopword removal, stemming, dan tokenizing agar data siap digunakan untuk analisis. Setelah itu, dilakukan ekstraksi fitur menggunakan teknik TF-IDF untuk memberikan bobot pada setiap kata dalam teks yang telah diproses. Selanjutnya, algoritma Naïve Bayes diterapkan untuk melakukan klasifikasi sentimen terhadap data tersebut, mengelompokkan opini menjadi kategori positif, negatif, atau netral.</p>			
4	ANALISIS SENTIMEN KAUM LGBT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN	Alvionita Mila Anjani, Ahmad Abdul Chamid, Alif Catur Murti	JURNAL TEKNIK INFORMATIKA, Vol.1 No.2 Juli 2022, Halaman 1~8 ISSN:2828-4445

	ALGORITMA <i>NAÏVE BAYES</i>		
	<p>Masalah Yang Ditemukan:</p> <p>Data mentah dari Twitter berupa tweet bersifat tidak terstruktur dan perlu diolah terlebih dahulu sebelum dapat digunakan dalam analisis sentimen. Selain itu, tweet sering kali mengandung noise seperti mention, hashtag, URL, dan tanda baca yang harus dihilangkan untuk meningkatkan kualitas data. Tantangan lain yang dihadapi adalah ketidakseimbangan data, di mana jumlah data pada setiap kelas sentimen tidak merata; makalah mencatat adanya polaritas positif sebesar 77% dan negatif sebesar 150,2%, yang menunjukkan ketidakseimbangan signifikan dalam distribusi kelas sentimen.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini melakukan beberapa tahapan preprocessing data, seperti case folding, cleansing, tokenizing, stopword removal, dan stemming, guna membersihkan serta menata data sebelum analisis dilakukan. Meskipun tidak dijelaskan secara rinci, kemungkinan besar teknik ekstraksi fitur yang digunakan adalah TF-IDF atau metode serupa untuk memberikan bobot pada kata-kata dalam teks. Untuk proses klasifikasi sentimen, digunakan algoritma Naïve Bayes yang dikenal memiliki tingkat akurasi tinggi dalam pengklasifikasian data teks.</p>		
5	Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan	Sella Adelia, Eva Milanda, Jena Santari, Dea Talia	Journal of Information Technology Ampera Vol. 4, No. 3, 2023 e-ISSN: 2774-2121

	Algoritma Klasifikasi <i>Naïve Bayes</i>	Kesuma, Echa Silvia, Fandi Kurniawan	
	<p>Masalah Yang Ditemukan:</p> <p>Pengumpulan data dilakukan dengan mengambil komentar YouTube yang relevan dan representatif, khususnya dari video tutorial pemrograman, meskipun komentar tersebut mungkin tidak sepenuhnya mencerminkan sentimen umum. Data mentah dari komentar ini mengandung noise seperti simbol, URL, dan kata-kata tidak relevan, sehingga perlu melalui tahap preprocessing yang mencakup pembersihan data, case folding, tokenizing, normalisasi, filtering, dan stemming. Selain itu, terdapat kemungkinan ketidakseimbangan data pada tiap kelas sentimen (positif, negatif, netral), yang dapat memengaruhi hasil analisis.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Pengumpulan data dilakukan melalui teknik web scraping untuk mengambil komentar dari video tutorial pemrograman di YouTube, meskipun kata kunci yang digunakan tidak dijelaskan secara eksplisit dalam makalah. Setelah data dikumpulkan, dilakukan tahapan preprocessing seperti yang telah dijelaskan sebelumnya guna membersihkan dan mempersiapkan data sebelum dianalisis. Untuk klasifikasi sentimen, penelitian ini menggunakan algoritma Naïve Bayes yang dikenal efektif dalam pengolahan teks. Evaluasi model, meskipun tidak dijelaskan secara rinci, kemungkinan menggunakan metrik seperti accuracy, precision, recall, dan F1-score, dengan hasil menunjukkan akurasi 100% untuk sentimen positif dan 66,39% untuk sentimen negatif.</p>		

6	Pengaruh Penggunaan Media Sosial Tiktok dalam Proses Pembelajaran di Sekolah Dasar Negeri Griya Bandung Indah	Renata Nur Annisa, Dinie Anggraeni Dewi, Santi Nurhayati	DIRASAH Volume 6, Number 2, Agustus 2023 p-ISSN: 2615-0212 e-ISSN: 2621-2838 https://ejournal.iaifa.ac.id/index.php/dirasaah
<p>Masalah Yang Ditemukan:</p> <p>Penggunaan TikTok yang berlebihan oleh siswa telah menyebabkan mereka menjadi lalai dalam belajar dan mengganggu proses pembelajaran di kelas, karena perhatian mereka lebih tertuju pada konten viral daripada materi pelajaran. Hal ini turut berdampak pada menurunnya minat belajar, karena siswa cenderung lebih tertarik pada hiburan yang disajikan di TikTok dibandingkan dengan mengikuti kegiatan pembelajaran secara aktif.</p>			
<p>Solusi Yang Dilakukan:</p> <p>Peneliti melakukan observasi langsung di kelas untuk mengamati secara nyata pengaruh penggunaan TikTok terhadap proses pembelajaran, sehingga memungkinkan pemahaman yang lebih kontekstual terhadap perilaku siswa. Selain itu, wawancara dengan guru dan siswa juga dilakukan guna menggali informasi yang lebih mendalam mengenai dampak TikTok terhadap minat dan konsentrasi belajar. Untuk memperkuat temuan, penelitian ini juga didukung oleh studi literatur yang merujuk pada berbagai sumber relevan tentang pengaruh media sosial terhadap proses pembelajaran, sehingga memberikan landasan teoritis yang kuat.</p>			
7	KOMPARASI METODE ALGORITMA KLASIFIKASI	Deny Setiawan Wijaya, Destriana	Jurnal TEKINKOM, Volume 7, Nomor 1, Juni 2024 DOI: 10.37600/tekinkom.v7i1.1392 E-ISSN: 2621-3079 ISSN: 2621-1556

<p>PADA ANALISIS SENTIMEN KOMENTAR CYBERBULL YING DI INSTAGRAM</p>	<p>Widyaning rum</p>	
<p>Masalah Yang Ditemukan:</p> <p>Penelitian ini bertujuan mengidentifikasi komentar yang mengandung unsur cyberbullying dari sejumlah besar komentar di Instagram, yang memerlukan penerapan teknik text mining dan analisis sentimen yang efektif. Untuk mengklasifikasikan sentimen komentar, dipilih algoritma klasifikasi yang tepat dengan membandingkan kinerja antara Naïve Bayes dan Decision Tree. Data mentah berupa komentar Instagram yang tidak terstruktur terlebih dahulu diproses melalui tahap pembersihan untuk menghilangkan noise seperti simbol, URL, dan tanda baca, agar siap digunakan dalam analisis lebih lanjut.</p>		
<p>Solusi Yang Dilakukan:</p> <p>Data dikumpulkan dari komentar akun Instagram artis Indonesia dalam rentang waktu September 2023 hingga Mei 2024, kemudian diproses menggunakan teknik text mining yang mencakup case folding, word repair, stemming, tokenizing, stopword removal, dan translate untuk membersihkan data dari noise dan mempersiapkannya bagi analisis sentimen. Dalam proses klasifikasi, dua algoritma yakni Naïve Bayes dan Decision Tree digunakan dan dibandingkan, dengan keduanya memanfaatkan teknik TF-IDF untuk ekstraksi fitur dari data teks yang telah diproses.</p>		

8	PENGARUH MEDIA SOSIAL BAGI PROSES BELAJAR SISWA	Kadek Widya Handayani	
<p>Masalah Yang Ditemukan:</p> <p>Penggunaan TikTok yang berlebihan di kalangan siswa telah menyebabkan mereka mengabaikan kegiatan belajar dan menurunnya minat terhadap materi pelajaran, karena konten viral di platform tersebut lebih menarik perhatian mereka. Secara umum, media sosial juga memberikan dampak negatif terhadap proses belajar, seperti menurunnya konsentrasi, munculnya kecanduan, serta dampak psikologis lainnya. Penelitian ini menggunakan metode survei, studi kasus, dan penelitian lapangan, namun masing-masing metode memiliki keterbatasan, terutama dalam hal generalisasi temuan, karena studi kasus hanya difokuskan pada satu sekolah dasar sehingga hasilnya sulit mewakili populasi secara luas.</p>			
<p>Solusi Yang Dilakukan:</p> <p>Peneliti melakukan observasi langsung di kelas untuk mengamati perilaku siswa dan memahami pengaruh penggunaan TikTok terhadap proses pembelajaran. Selain itu, wawancara dengan guru dan siswa kemungkinan dilakukan guna memperoleh informasi yang lebih mendalam mengenai dampak tersebut. Makalah ini juga memberikan saran untuk memanfaatkan media sosial secara positif dalam pembelajaran, seperti melalui pembentukan komunitas belajar, berbagi sumber belajar, serta meningkatkan interaksi antara guru dan siswa guna menciptakan lingkungan belajar yang lebih produktif dan kolaboratif.</p>			

9	Dampak Penggunaan Media Sosial terhadap Interaksi Sosial Masyarakat: Studi Literature	Nabila Diva Pratidina, Jane Mitha	Jurnal Ilmiah Universitas Batanghari Jambi 23, 1 (2023): 810-815 DOI: 10.33087/jiubj.v23i1.3083 http://ji.unbari.ac.id/index.php/ilmiah
<p>Masalah Yang Ditemukan:</p> <p>Media sosial memiliki dampak positif dan negatif terhadap interaksi sosial; di satu sisi, media sosial dapat meningkatkan kerjasama, mengurangi pertikaian, dan memperluas jaringan pertemanan, namun di sisi lain, juga dapat menjauhkan hubungan personal, mengurangi interaksi tatap muka, dan memicu peningkatan konflik. Penelitian ini menggunakan metode literature review, sehingga analisis sepenuhnya didasarkan pada literatur yang ada tanpa pengumpulan data primer, yang membatasi generalisasi temuan dan membuka kemungkinan adanya bias dalam interpretasi. Meskipun makalah berhasil mengidentifikasi berbagai dampak negatif media sosial, sayangnya belum menawarkan solusi yang komprehensif untuk mengatasinya.</p>			
<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini melakukan literature review yang komprehensif dengan menggunakan kata kunci yang relevan untuk menelusuri literatur terkait dampak media sosial terhadap interaksi sosial. Temuan dari berbagai sumber dianalisis untuk mengidentifikasi dampak positif, seperti peningkatan kerjasama dan perluasan jaringan sosial, serta dampak negatif seperti penurunan interaksi tatap muka dan meningkatnya konflik. Sebagai respons terhadap temuan tersebut, penelitian ini menyarankan beberapa solusi untuk mengurangi dampak negatif media sosial, antara lain melalui pengawasan penggunaan media sosial pada anak, penggunaan yang bijak</p>			

	dan bertanggung jawab, serta pemberian bimbingan konseling jika diperlukan.		
10	Analisis Sentimen Komentar Youtube terhadap Food Vlogger dengan Menggunakan Metode <i>Naïve Bayes</i>	Sulistia Maharani Harahap, Rakhamat Kurniawan	Volume 9 No. 1, Juni 2024 http://ejournal.ust.ac.id/index.php/Jurnal_Means/ MEANS (Media Informasi Analisa dan Sistem) p-ISSN: 2548-6985, e-ISSN:2599-3089
	<p>Masalah Yang Ditemukan:</p> <p>Pengumpulan data dilakukan dengan mengambil komentar YouTube yang relevan dan representatif untuk analisis sentimen, meskipun data tersebut mungkin tidak sepenuhnya mencerminkan sentimen publik secara keseluruhan terhadap food vlogger. Data mentah kemudian diproses melalui tahap preprocessing yang mencakup cleaning, case folding, tokenizing, stopword removal, stemming, dan normalization guna memastikan data siap untuk dianalisis. Penelitian ini menghadapi tantangan imbalance data, di mana jumlah data pada tiap kelas sentimen tidak seimbang, yang tercermin dari hasil evaluasi model dengan recall mencapai 100% namun precision hanya 16%. Akurasi model juga tergolong rendah, yaitu sebesar 44,8276%, yang menunjukkan perlunya peningkatan dalam pemodelan atau penanganan ketidakseimbangan data.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini mengumpulkan data komentar YouTube yang relevan dan representatif untuk analisis sentimen, meskipun data yang diperoleh mungkin tidak sepenuhnya mencerminkan sentimen publik secara keseluruhan terhadap food vlogger. Data mentah kemudian diproses</p>		

	<p>melalui tahapan preprocessing, yang meliputi cleaning, case folding, tokenizing, stopword removal, stemming, dan normalization, agar sesuai untuk dianalisis. Salah satu tantangan yang dihadapi adalah ketidakseimbangan data (imbalance data) antar kelas sentimen, yang ditunjukkan dengan recall mencapai 100% namun precision hanya 16%, mencerminkan adanya bias klasifikasi. Meskipun model diharapkan memiliki akurasi tinggi, hasil menunjukkan akurasi yang relatif rendah sebesar 44,8276%, sehingga diperlukan perbaikan lebih lanjut dalam pengolahan data atau pemilihan model.</p>		
1	Analisis Sentimen Pendapat Masyarakat Terhadap PPKM DKI Jakarta Dengan Metode <i>Naïve Bayes</i>	Hikari Ardhiansya, Auliya Burhanuddin, Shintia Dwi Alika	60 Jurnal Teknik Informatika dan Sistem Informasi Vol. 10, No. 2, Juni 2023, Hal. 60-70 ISSN 2407-4322 E-ISSN 2503-2933
	<p>Masalah Yang Ditemukan:</p> <p>Pengumpulan data dilakukan dengan mengambil komentar dari konten berita tertentu di YouTube yang dianggap relevan dan representatif untuk analisis sentimen, meskipun demikian, data tersebut mungkin tidak sepenuhnya mencerminkan sentimen masyarakat secara keseluruhan. Penelitian ini juga menghadapi tantangan imbalance data, di mana distribusi komentar tidak merata antar kelas sentimen, dengan dominasi sentimen netral sebesar 87,2%. Meskipun demikian, model klasifikasi sentimen yang digunakan mampu mencapai akurasi yang cukup tinggi, yaitu 90% untuk klasifikasi dua kelas dan 81% untuk tiga kelas, menunjukkan performa yang cukup baik meskipun ada ketidakseimbangan data.</p>		

	<p>Solusi Yang Dilakukan:</p> <p>Pengumpulan data dilakukan melalui teknik web scraping dari komentar pada konten berita YouTube yang membahas PPKM di DKI Jakarta. Untuk mengklasifikasikan sentimen komentar, digunakan algoritma Naïve Bayes yang dikenal efektif dalam analisis teks. Evaluasi model dilakukan menggunakan confusion matrix dengan mengukur kinerja melalui metrik akurasi, presisi, recall, dan F1-score. Dua model klasifikasi diuji, yakni model dua kelas (positif dan negatif) serta model tiga kelas (positif, netral, dan negatif), guna membandingkan efektivitas klasifikasi dalam skenario berbeda.</p>		
1 2	<p>KLASIFIKASI DATA TWEET UJARAN KEBENCIAN DI MEDIA SOSIAL MENGGUNAK AN <i>NAÏVE</i> <i>BAYES</i> CLASSIFIER</p>	<p>Noor Aliyah Susanti, Miftahul Walid, Hoiriyah</p>	<p>Vol. 6 No. 2, September 2022 JATI (Jurnal Mahasiswa Teknik Informatika)</p>
	<p>Masalah Yang Ditemukan:</p> <p>Banyak penelitian menggunakan data mentah dari media sosial seperti Twitter, YouTube, dan Instagram, yang bersifat tidak terstruktur dan memerlukan proses pembersihan serta pemrosesan sebelum dapat dianalisis. Data ini sering kali mengandung noise berupa mention, hashtag, URL, dan tanda baca yang tidak relevan dan harus dihilangkan agar tidak mengganggu hasil analisis. Selain itu, tantangan lain yang umum dihadapi adalah imbalance data, di mana jumlah data pada setiap kelas sentimen (positif, negatif, netral) tidak seimbang, sehingga dapat memengaruhi akurasi model. Meskipun mencapai akurasi tinggi dalam klasifikasi</p>		

	sentimen menjadi tujuan utama, hal ini sering kali sulit dicapai akibat kompleksitas data yang dihadapi.		
	<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini melibatkan berbagai tahapan preprocessing data, seperti case folding, cleaning, tokenizing, stopword removal, stemming, dan normalization, untuk membersihkan dan menata data sebelum dianalisis. Dalam proses klasifikasi sentimen, digunakan algoritma seperti Naïve Bayes, Decision Tree, atau metode lain yang sesuai guna mengelompokkan data ke dalam kategori sentimen tertentu. Kinerja model kemudian dievaluasi menggunakan confusion matrix dan metrik seperti akurasi, presisi, recall, dan F1-score. Seluruh proses, mulai dari pengumpulan data hingga klasifikasi, didukung oleh penggunaan tools seperti RapidMiner, Python, dan berbagai library terkait yang memfasilitasi pemrosesan data secara efisien.</p>		
1 3	ANALISIS SENTIMEN TERHADAP BERHENTINYA TIKTOKSHOP PADA MEDIA SOSIAL TWITTER MENGUNAKAN ALGORITMA NAÏVE BAYES	Riki Sanjaya, Edi Tohidi, Edi Wahyudi, Kaslan	JATI (Jurnal Mahasiswa Teknik Informatika) Vol. 8 No. 1, Februari 2024

	<p>Masalah Yang Ditemukan:</p> <ul style="list-style-type: none"> • Data tidak terstruktur dan mengandung noise: Data mentah dari media sosial (YouTube, Twitter) mengandung noise (misalnya, mention, hashtag, URL, tanda baca) dan tidak terstruktur, sehingga perlu dibersihkan dan diproses sebelum analisis. • Pemilihan fitur: Memilih fitur yang tepat untuk mewakili sentimen dalam data teks merupakan tantangan. Beberapa penelitian menggunakan TF-IDF, tetapi tidak selalu dijelaskan secara detail bagaimana fitur dipilih. • Imbalance data: Jumlah data untuk setiap kelas sentimen (positif, negatif, netral) seringkali tidak seimbang, yang dapat mempengaruhi hasil analisis. Beberapa penelitian menunjukkan ketidakseimbangan yang signifikan. 		
	<p>Solusi Yang Dilakukan:</p> <p>Berbagai tahapan preprocessing dilakukan untuk membersihkan dan menata data, termasuk case folding, cleaning, tokenizing, stopword removal, stemming, dan normalization, agar data siap untuk dianalisis. Setelah itu, fitur-fitur penting dari data teks diekstraksi menggunakan teknik TF-IDF untuk memberikan bobot pada setiap kata. Dalam proses klasifikasi sentimen, algoritma Naïve Bayes digunakan secara luas karena kesederhanaan dan efektivitasnya, meskipun beberapa penelitian juga menerapkan algoritma lain seperti Decision Tree, Random Forest, atau Support Vector Machine untuk membandingkan performa dan meningkatkan akurasi klasifikasi.</p>		
1 4	ANALISIS SENTIMEN MASYARAKA T TERHADAP PAYLATER	Alfandi Safira, Firman Noor Hasan	ZONAsi Alfandi Safira, Analisis Sentimen Masyarakat...v ISSN: 1978-1520 ISSN:2656-7407 2656-7393 ZONAsi VOL.5NO.1 Page: 59-70 JANUARI 2023

	MENGGUNAKAN METODE <i>NAÏVE BAYES</i> CLASSIFIER		
	<p>Masalah Yang Ditemukan:</p> <p>Data mentah dari Twitter berupa tweet bersifat tidak terstruktur dan perlu dibersihkan serta diproses terlebih dahulu sebelum dapat digunakan untuk analisis sentimen. Tweet juga mengandung berbagai noise seperti mention, hashtag, dan URL yang harus dihilangkan agar tidak mengganggu hasil analisis. Selain itu, pengumpulan data menggunakan fitur advanced search di Twitter dan pustaka sncrape dapat membatasi jumlah data yang berhasil dikumpulkan, seperti dalam penelitian ini yang hanya memperoleh sebanyak 405 data tweet, sehingga dapat memengaruhi representativitas dan keakuratan hasil analisis.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Berbagai tahapan preprocessing dilakukan untuk membersihkan dan menata data, termasuk case folding, cleaning, tokenizing, stopword removal, stemming, dan normalization, agar data siap untuk dianalisis. Setelah itu, fitur-fitur penting dari data teks diekstraksi menggunakan teknik TF-IDF untuk memberikan bobot pada setiap kata. Dalam proses klasifikasi sentimen, algoritma Naïve Bayes digunakan secara luas karena kesederhanaan dan efektivitasnya; meskipun demikian, beberapa penelitian juga menerapkan algoritma lain seperti Decision Tree, Random Forest, atau Support Vector Machine guna membandingkan performa dan meningkatkan akurasi klasifikasi.</p>		
1 5	Analisis Sentimen Gofood Berdasarkan Twitter	Melati Indah Petiwi, Agung Triayudi,	JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 6, Nomor 1, Januari 2022, Page 542-550 ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at

	Menggunakan Metode <i>Naïve Bayes</i> dan Support Vector Machine	Ira Diana Sholihati	https://ejurnal.stmik-budidarma.ac.id/index.php/mib
	<p>Masalah Yang Ditemukan:</p> <p>Data mentah dari Twitter berupa tweet bersifat tidak terstruktur, sehingga perlu dibersihkan dan diproses terlebih dahulu sebelum dilakukan analisis sentimen. Tweet tersebut juga mengandung berbagai noise seperti mention, hashtag, URL, emoji, dan tanda baca yang harus dihilangkan untuk meningkatkan kualitas data. Selain itu, terdapat masalah imbalance data, di mana jumlah data pada setiap kelas sentimen (positif, negatif, netral) kemungkinan tidak seimbang, yang dapat memengaruhi akurasi dan validitas hasil analisis; dalam penelitian ini, ditemukan dominasi sentimen netral sebesar 92,8%.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Data dalam penelitian ini diproses melalui beberapa tahapan preprocessing, yaitu case folding, filtering, tokenizing, dan stopword removal, guna membersihkan dan menyiapkan data untuk analisis. Pengumpulan data dilakukan menggunakan library Tweepy dan Twitter API dengan kata kunci "Gofood" untuk memperoleh tweet yang relevan. Dalam analisis sentimen, digunakan dua algoritma klasifikasi, yaitu <i>Naïve Bayes</i> dan Support Vector Machine (SVM), yang hasilnya kemudian dibandingkan untuk menilai efektivitas masing-masing metode</p>		
1 6	Analisis Sentimen Masyarakat Berdasarkan Komentar Kerja Sama Tiktok	Zaenab Kurnia	submission date: 06-Jun-2024 04:13PM (UTC+0700) Submission ID: 2396781313

	Shop dan Tokopediadi Instagram Menggunakan Metode <i>Naive Bayes</i> Classifier		File name: JPTIS_Vol_2_no_2_Juni_2024_hal_115-125.pdf (945.96K) Word count: 3084 Character count: 19487
	<p>Masalah Yang Ditemukan:</p> <p>Data mentah berupa komentar Instagram bersifat tidak terstruktur dan perlu dibersihkan terlebih dahulu sebelum dapat digunakan untuk analisis sentimen. Komentar tersebut mengandung berbagai noise seperti mention, hashtag, URL, emoji, dan tanda baca yang harus dihilangkan untuk memastikan kualitas data yang baik. Selain itu, terdapat masalah ketidakseimbangan data (imbalance data), di mana jumlah data pada setiap kelas sentimen (positif, negatif, netral) tidak seimbang, dan sebagian besar data bahkan tidak terklasifikasi, yakni 437 dari total 641 komentar, yang dapat memengaruhi keakuratan dan validitas hasil analisis.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Data dalam penelitian ini diproses melalui beberapa tahapan preprocessing, yaitu case folding, tokenizing, stopword removal, dan stemming untuk membersihkan serta menyiapkan data sebelum analisis. Pengumpulan data dilakukan dari komentar pada dua postingan Instagram, masing-masing berasal dari akun resmi TikTok dan Tokopedia, yang membahas kolaborasi antara kedua platform tersebut. Selanjutnya, data diberi label sentimen (positif, negatif, netral) oleh seorang pakar guna memastikan akurasi dalam proses klasifikasi.</p>		
1 7	Analisis Sentimen Cyberbullying pada Media Sosial Twitter	Rusiana Hilma, Munirul Ula, Sayed	TECHSI: Vol. 14, Issue. 2, 2023

	menggunakan Metode Support Vector Machine dan <i>Naïve Bayes</i> Classifier	Fachrurrazi	
	<p>Masalah Yang Ditemukan:</p> <p>Data mentah dari Twitter berupa tweet bersifat tidak terstruktur dan harus dibersihkan terlebih dahulu sebelum dapat digunakan untuk analisis sentimen. Tweet tersebut umumnya mengandung berbagai noise seperti mention, hashtag, URL, emoji, dan tanda baca yang perlu dihilangkan agar tidak mengganggu hasil analisis. Selain itu, terdapat kemungkinan terjadinya imbalance data, yaitu ketidakseimbangan jumlah data pada setiap kelas sentimen (positif, negatif, netral), meskipun hal ini tidak dijelaskan secara eksplisit dalam makalah.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Data dalam penelitian ini diproses melalui beberapa tahapan preprocessing, termasuk case folding, normalisasi kata, stopword removal, tokenizing, dan stemming, untuk membersihkan dan menyiapkan data sebelum dianalisis. Pengumpulan data dilakukan dari Twitter menggunakan Twitter API dengan memasukkan 10 kata kunci yang berpotensi menimbulkan cyberbullying guna mendapatkan tweet yang relevan. Kinerja model dalam analisis sentimen kemudian dievaluasi menggunakan metrik akurasi untuk mengukur tingkat keberhasilan klasifikasi.</p>		
18	Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone	Kairil Anwar	<p>KLIK: Kajian Ilmiah Informatika dan Komputer ISSN 2723-3898 (Media Online) Vol 2, No 4, Februari 2022 Hal 148-155 https://djournals.com/klik</p>

	Menggunakan <i>Naïve Bayes</i>		
<p>Masalah Yang Ditemukan:</p> <p>Data mentah berupa komentar Instagram pada video review smartphone bersifat tidak terstruktur dan perlu dibersihkan terlebih dahulu sebelum dilakukan analisis sentimen. Komentar-komentar tersebut mengandung berbagai noise seperti mention, hashtag, URL, emoji, dan tanda baca yang harus dihilangkan agar tidak mengganggu kualitas analisis. Tujuan utama dari penelitian ini adalah mencapai akurasi tinggi dalam klasifikasi sentimen, namun hasil yang dilaporkan menunjukkan variasi akurasi, yaitu sebesar 73% untuk algoritma Gaussian Naïve Bayes dan 83% untuk Multinomial Naïve Bayes.</p>			
<p>Solusi Yang Dilakukan:</p> <p>Data dalam penelitian ini diproses melalui beberapa tahapan preprocessing, yaitu case folding, tokenizing, dan filtering, untuk membersihkan dan menyiapkan data sebelum dianalisis. Pengumpulan data dilakukan dengan metode scraping dari komentar pada video review smartphone di Instagram. Setelah itu, data diberi label sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral, guna mendukung proses klasifikasi sentimen secara akurat.</p>			
1 9	Implementasi Algoritma <i>Naïve Bayes</i> Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada	Raihan Putra Mohamma d Rosidi, Kiki Setiawan	Jurnal Indonesia : Manajemen Informatika dan Komunikasi www.journal.amikindonesia.ac.i Vol 5 No 1, Januari (2024) E-ISSN: 2723-7079, P- ISSN: 2776-8074d/jimik/

	<p>Masalah Yang Ditemukan:</p> <p>Data penjualan kantin sebelumnya tidak tersusun rapi dan hanya disimpan sebagai arsip tanpa dimanfaatkan untuk analisis, sehingga menghambat pemahaman terhadap perilaku konsumen dan pengoptimalan strategi penjualan. Selain itu, metode analisis data tradisional dianggap tidak memadai untuk memahami kompleksitas perilaku konsumen di kantin, sehingga informasi yang diperlukan untuk meningkatkan strategi penjualan dan kepuasan pelanggan menjadi kurang tersedia.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>Penelitian ini menerapkan algoritma Naïve Bayes sebagai solusi untuk menganalisis data penjualan dan mengidentifikasi pola pembelian konsumen, dengan mempertimbangkan kemampuannya dalam mengklasifikasikan data berdasarkan probabilitas dan pengalaman sebelumnya. Data penjualan dikumpulkan dan melalui tahapan pra-pemrosesan untuk memastikan format dan struktur data sesuai dengan kebutuhan algoritma. Atribut-atribut relevan seperti jenis produk, tanggal pembelian, jumlah pembelian, dan harga disusun secara sistematis guna mendukung proses analisis yang akurat dan bermakna.</p>		
20	<p>Implementation of RFM Analysis to Enhance Sales Patterns of Food and Beverages at Bonjour Café and Resto Using the Apriori Algorithm</p>	<p>Julvan Marzuki Putra Sibarani, Yuma Akbar, Sutisna, Kiki Setiawan</p>	<p>International Journal Software Engineering and Computer Science (IJSECS) 4 (3), 2024, 1261-1279 Published Online December 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs) P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i3.3073.</p>

	<p>Masalah Yang Ditemukan:</p> <p>Intense Competition in the Culinary Business: The rapid growth of the culinary business has led to intense competition. Bonjour Cafe Resto, despite having a wide variety of menu options, needs to innovate to increase customer satisfaction and sales.</p>		
	<p>Solusi Yang Dilakukan:</p> <p>RFM Analysis and Apriori Algorithm: This research utilizes RFM (Recency, Frequency, Monetary) analysis to analyze customer behavior based on sales transaction data. RFM attributes (last transaction date, transaction frequency, and total transaction value) are used to segment customers into various groups. The Apriori algorithm is employed to identify frequently purchased itemsets and generate association rules to recommend menu combinations.</p>		
21	<p>PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus</p>	<p>Gellysa Urva, dkk.</p>	<p>PT. Sonpedia Publishing Indonesia</p>
	<p>Masalah Yang Ditemukan:</p> <p>Analisis sentimen memiliki keterbatasan, terutama karena banyak metode yang belum mampu menangkap nuansa emosi dalam teks, seperti sarkasme atau konteks budaya yang tersembunyi, sehingga dapat menyebabkan kesalahan interpretasi. Selain itu, kualitas data yang digunakan sering kali buruk, berupa data yang tidak bersih atau tidak terstruktur, yang pada akhirnya dapat memengaruhi akurasi dan keandalan model dalam mengklasifikasikan sentimen secara tepat.</p>		

	<p>Solusi Yang Dilakukan:</p> <p>Peningkatan preprocessing data dilakukan melalui proses pembersihan dan normalisasi untuk memastikan kualitas input yang lebih baik ke dalam model analisis sentimen, sehingga hasil klasifikasi menjadi lebih akurat dan andal. Selain itu, penggunaan teknik pembelajaran mesin lain seperti menggabungkan algoritma Naïve Bayes dengan metode lain seperti Support Vector Machines (SVM) atau deep learning juga dapat diterapkan guna meningkatkan performa dan akurasi model dalam memahami serta mengklasifikasikan sentimen secara lebih efektif.</p>		
22	DATA SCIENCE GEOSPASIAL	Yulia Fransisca A	Penerbit Andi
	<p>Masalah Yang Ditemukan:</p> <p>Penggunaan data geospasial dalam analisis masih menghadapi keterbatasan, di mana banyak studi belum memanfaatkannya secara optimal, sehingga informasi yang dihasilkan cenderung kurang akurat dan kurang mencerminkan kondisi spasial sebenarnya. Selain itu, integrasi antara data geospasial dan data non-geospasial sering kali menjadi tantangan tersendiri karena perbedaan format, struktur, dan skala data, yang dapat menghambat terciptanya analisis yang komprehensif dan menyeluruh.</p> <p>Solusi Yang Dilakukan:</p> <p>Peningkatan pelatihan dan penyediaan sumber daya edukasi sangat penting untuk meningkatkan pemahaman dan keterampilan dalam mengelola serta menganalisis data geospasial secara efektif. Selain itu, pengembangan alat integrasi yang mampu menggabungkan data geospasial dan non-geospasial secara efisien juga diperlukan guna mendukung analisis yang lebih holistik dan komprehensif, sehingga informasi yang dihasilkan menjadi lebih akurat dan bermanfaat dalam pengambilan Keputusan.</p>		

23	<p>PENERAPAN METODE <i>NAÏVE</i> <i>BAYES</i> DNA SKALA LIKERT PADA APLIKASI PREDIKSI KELULUS MAHASISWA</p>	<p>Fadila Woro Isti Rahayu, S. T., M.T., M. Harry K. Saputra, S.T., M.T.I.</p>	<p>Kreatif Industri Nusantara</p>
<p>Masalah Yang Ditemukan:</p> <p>Metode Naïve Bayes sering kali memiliki keterbatasan dalam hal akurasi, terutama ketika diterapkan pada dataset yang kompleks, di mana performanya cenderung lebih rendah dibandingkan dengan algoritma pembelajaran mesin lainnya. Salah satu kelemahan utama dari algoritma ini adalah asumsi independensi antar fitur, yang dalam praktik jarang terpenuhi, sehingga dapat mengakibatkan kesalahan dalam prediksi dan menurunkan keandalan hasil analisis.</p> <p>Solusi Yang Dilakukan:</p> <p>Untuk meningkatkan akurasi model, Naïve Bayes dapat dikombinasikan dengan teknik ensemble seperti Random Forest atau Gradient Boosting, yang mampu mengurangi kelemahan model tunggal dan menghasilkan prediksi yang lebih andal. Selain itu, penerapan teknik preprocessing yang lebih baik, seperti normalisasi dan pengurangan dimensi, juga penting untuk meningkatkan kualitas data yang digunakan, sehingga model dapat bekerja secara lebih efektif dalam menganalisis dan mengklasifikasikan data secara akurat.</p>			
24	<p>KLASIFIKASI KONDISI PERALATAN ELEKTRONIK</p>	<p>Ayu Lestari Nasution, Rd Nuraini Siti Fathonah</p>	<p>Penerbit Buku Pedia</p>

	METODE GAUSSIAN <i>NAÏVE BAYES</i>		
	<p>Masalah Yang Ditemukan:</p> <p>Metode Gaussian Naïve Bayes memiliki keterbatasan dalam hal akurasi klasifikasi, khususnya ketika diterapkan untuk mengklasifikasikan kondisi peralatan elektronik, karena performanya dapat menurun jika data tidak mengikuti distribusi normal. Hal ini disebabkan oleh asumsi dasar algoritma tersebut yang menganggap bahwa setiap fitur memiliki distribusi normal, padahal dalam praktiknya, asumsi ini sering kali tidak terpenuhi, sehingga dapat mengakibatkan kesalahan dalam prediksi dan menurunkan keandalan hasil analisis.</p> <p>Solusi Yang Dilakukan:</p> <p>Penerapan teknik preprocessing yang lebih baik, seperti pembersihan data dan normalisasi, sangat penting untuk mengurangi pengaruh outlier dan memastikan data lebih sesuai dengan asumsi distribusi yang dibutuhkan oleh model, seperti Gaussian Naïve Bayes. Selain itu, penggunaan model hibrida dengan menggabungkan Naïve Bayes bersama algoritma lain seperti Decision Trees atau Random Forest dapat meningkatkan akurasi klasifikasi, karena pendekatan ini memanfaatkan keunggulan masing-masing metode untuk menghasilkan prediksi yang lebih andal dan akurat.</p>		
25	ANALISIS SENTIMENT MASYARAKAT TERHADAP KEBIJAKAN POLISI TILANG MANUAL DI INDONESIA DENGAN SVM (Muhammad Syarul Amrullah, Syafrial Fachri Pane	Penerbit Buku Pedia

	SUPPORT VECTOR MACHINE)		
<p>Masalah Yang Ditemukan:</p> <p>Masyarakat menunjukkan resistensi terhadap kebijakan tilang manual yang diterapkan oleh polisi, yang sering kali memicu ketidakpuasan hingga protes atau penolakan. Salah satu penyebab utama dari penolakan ini adalah kurangnya pemahaman masyarakat mengenai tujuan dan manfaat dari kebijakan tersebut, sehingga menimbulkan kesalahpahaman yang berdampak negatif terhadap penerimaan dan efektivitas pelaksanaan kebijakan di lapangan.</p> <p>Solusi Yang Dilakukan:</p> <p>Untuk meningkatkan penerimaan masyarakat terhadap kebijakan tilang manual, diperlukan kampanye edukasi dan sosialisasi yang efektif guna menjelaskan tujuan dan manfaat dari kebijakan tersebut, sehingga masyarakat dapat lebih memahami dan menerimanya dengan baik. Selain itu, penggunaan metode analisis sentimen yang lebih canggih, seperti pemanfaatan survei dan wawancara, dapat memberikan gambaran yang lebih akurat mengenai opini publik, sehingga kebijakan yang diterapkan dapat disesuaikan dengan kondisi dan persepsi masyarakat secara lebih tepat.</p>			

2.1.4 Matrik Perbandingan RP,RQ, dan RO

Masalah Penelitian, Identifikasi Masalah dan Tujuan Penelitian Dari hasil tinjauan pustaka terkait judul proposal yang di tulis dan di rangkumsecara keseluruhan adalah sebagai berikut yaitu :

tabel 2.4 Matrix Perbandingan RP,RQ, dan RO

Identifikasi Masalah (Research Problem / RP)		Rumusan Masalah (Research Question / RQ)		Tujuan Penelitian (Research Objective / RO)	
RP-1	Terdapat peningkatan signifikan dalam penggunaan media sosial, khususnya Instagram, oleh anak-anak di bawah usia 13 tahun, yang melanggar batasan usia yang telah ditetapkan oleh platform tersebut.	RQ-1	Faktor-faktor apa yang mempengaruhi penggunaan Instagram oleh anak usia dini di Indonesia?	RO-1	Menganalisis Penggunaan Instagram oleh Anak Usia Dini: Untuk memahami faktor-faktor yang mempengaruhi peningkatan penggunaan Instagram di kalangan anak-anak di bawah usia 13 tahun.
		RQ-2	Bagaimana pandangan dan opini masyarakat tentang dampak penggunaan Instagram pada anak di bawah 13 tahun?	RO-2	Mengidentifikasi Opini Publik: Untuk memetakan dan mengidentifikasi pandangan serta opini masyarakat mengenai dampak penggunaan Instagram oleh anak usia dini,

					baik yang positif maupun negatif.
RP- 2	Penggunaan Instagram oleh anak-anak dapat menimbulkan berbagai dampak negatif, termasuk paparan konten yang tidak sesuai, risiko cyberbullying, serta gangguan dalam perkembangan sosial mereka.	RQ- 3	Jenis konten apa yang diakses anak-anak di Instagram, dan bagaimana dampaknya pada perkembangan sosial dan psikologis mereka?	RO- 3	Mengedukasi Orang Tua dan Pendidik: Untuk menyediakan platform edukatif yang membantu orang tua dan pendidik memahami penggunaan media sosial oleh anak-anak dan memberikan panduan dalam pengawasan dan pendampingan mereka.
3.	Masyarakat, termasuk orang tua dan pendidik, mungkin tidak memiliki pemahaman yang cukup mengenai bagaimana mengawasi dan mendampingi anak-anak dalam				

	menggunakan media sosial secara sehat dan bertanggung jawab.				
--	--	--	--	--	--

2.2 Definisi dan Pengertian

2.2.1 Anak usia dini menggunakan media sosial

Periode ini merupakan masa perkembangan yang sangat pesat, baik fisik, kognitif, maupun sosial-emosional. Penggunaan media sosial oleh anak usia dini merupakan fenomena kompleks yang melibatkan akses dan interaksi anak dengan platform digital seperti YouTube, Instagram, TikTok, dan lain-lain. Penggunaan ini bisa bervariasi, mulai dari sekadar menonton video hingga berinteraksi aktif dengan konten dan pengguna lain. Penting untuk diingat bahwa akses dan penggunaan ini seringkali terjadi dengan tingkat pengawasan orangtua yang berbeda-beda.[13]

2.2.2 Analisis Sentimen

Sebuah metode dalam bidang *Natural Language Processing* (NLP), data mining, dan kecerdasan buatan yang bertujuan untuk mengidentifikasi, mengekstrak, serta mengklasifikasikan opini, emosi, atau sikap seseorang terhadap suatu entitas, topik, atau peristiwa tertentu berdasarkan data teks. Dengan kata lain, analisis sentimen berfungsi untuk memahami apakah suatu pernyataan mengandung sentimen positif, negatif, atau netral. Teknik analisis sentimen umumnya digunakan untuk menganalisis data yang berasal dari media sosial, ulasan produk, komentar, hingga berita online. Proses ini melibatkan beberapa tahapan, seperti pengumpulan data, pembersihan teks (*text preprocessing*), ekstraksi fitur, dan klasifikasi menggunakan algoritma pembelajaran mesin, salah satunya *Naïve Bayes*. Analisis sentimen digunakan untuk memetakan opini masyarakat mengenai

penggunaan Instagram oleh anak usia dini. Hasil analisis tidak hanya memberikan gambaran umum tentang persepsi publik, tetapi juga dapat menjadi dasar dalam perumusan kebijakan, strategi edukasi digital, serta pengembangan sistem informasi yang membantu orang tua dan pendidik dalam mengarahkan penggunaan media sosial secara sehat dan bertanggung jawab.

2.2.3 Instagram

Instagram adalah layanan jejaring sosial berbasis foto dan video yang memungkinkan pengguna untuk berbagi konten multimedia, berinteraksi dengan teman dan pengikut, serta mengikuti akun-akun yang menarik minat mereka. Pengguna dapat mengunggah foto dan video, mengeditnya dengan berbagai filter dan efek, dan menambahkan keterangan serta tagar untuk meningkatkan visibilitas konten mereka. Instagram tersedia di berbagai perangkat, termasuk smartphone (iOS dan Android), komputer, dan web. Platform ini digunakan secara luas oleh individu dan bisnis untuk berbagi momen pribadi, mempromosikan merek, dan membangun komunitas online. Namun, penting untuk menggunakan Instagram secara bertanggung jawab dan menyadari potensi dampak negatif, seperti kecanduan dan cyberbullying [14]. Selain sebagai media hiburan, Instagram juga berfungsi sebagai sarana komunikasi, promosi, bahkan pembelajaran. Fitur – fitur seperti *Stories*, *Reels*, *Live Streaming*, dan *Direct Message* memberikan pengalaman interaktif yang membuat pengguna semakin aktif berpartisipasi di dalamnya.

2.2.4 Naïve Bayes

Naïve Bayes adalah sebuah metode klasifikasi dalam pembelajaran mesin yang didasarkan pada penerapan Teorema Bayes dengan asumsi "naive" bahwa setiap fitur (atribut) dalam dataset berkontribusi secara independen terhadap kemungkinan suatu kelas. Meskipun asumsi independensi ini jarang benar dalam kenyataan, *Naïve Bayes* sering memberikan hasil yang baik dalam berbagai

aplikasi praktis.[15]

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

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X

P(H) : Probabilitas hipotesis H

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis

H

P(X) : Probabilitas X

Algoritma ini bekerja dengan menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan distribusi fitur yang ada. Sifat probabilistiknya membuat metode ini sederhana, cepat dan efisien untuk di aplikasikan pada dataset dengan ukuran besar maupun kecil. Salah satu keunggulan utama naive bayes adalah kemampuannya mengatasi masalah klasifikasi teks, seperti analisis sentimen, filtering spam email, hingga rekomendasi produk. Kelebihan lain dari algoritma naive bayes kemampuannya untuk tetap bekerja dengan baik meskipun terdapat keterbatasan data, serta hasil klasifikasinya yang relatif mudah diinterpretasikan. Namun demikian, metode ini juga memiliki keterbatasan, khususnya ketika akurasi prediksi. Oleh karena itu, dalam implementasinya diperlukan tahapan *preprocessing* data yang baik agar hasil klasifikasi menjadi lebih optimal.

2.2.5 Data

Dalam ilmu komputer dan pembelajaran mesin, dataset (kumpulan data) adalah sekumpulan data yang biasanya disusun dalam format tabel di mana setiap baris mewakili satu entitas atau rekaman, dan setiap kolom mewakili fitur atau atribut dari entitas tersebut. Dataset digunakan untuk berbagai tujuan, termasuk analisis data, pelatihan model pembelajaran mesin, dan pengujian algoritma.[3]

2.2.6 Rapid Minner

Sebuah platform perangkat lunak yang dirancang untuk analisis data, pembelajaran mesin, dan penambangan data (data mining). Platform ini menyediakan alat yang komprehensif untuk semua tahap proses analisis data, mulai dari pengumpulan data, preprocessing, pemodelan, evaluasi hingga implementasi model prediktif. RapidMiner populer karena kemudahan penggunaannya, antarmuka yang intuitif, serta dukungannya terhadap berbagai algoritma pembelajaran mesin [16]. Sebagai salah satu tools data mining *open-source* yang banyak digunakan, RapidMinner mendukung berbagai format data dan dapat diintegrasikan dengan bahasa pemrograman lain seperti Python dan R. Hal ini membuat RapidMiner fleksibel digunakan baik untuk penelitian akademis maupun kebutuhan industri. Fitur *drag-and-drop* yang dimilikinya mempermudah pengguna, bahkan bagi pemula yang tidak memiliki latar belakang pemrograman yang kuat, untuk menganalisis data yang kompleks. Dalam konteks penelitian ini, RapidMiner digunakan untuk mendukung proses analisis sentimen dengan algoritma naive bayes, mulai dari tahapan *cleasing data*, *tokenizing*, hingga klasifikasi sentimen ke dalam katagori positif, negatif, dan netral. Dengan bantuan RapidMiner, proses analisis dapat dilakukan lebih cepat, terstruktur dan mudah divisualisirkan, sehingga hasil penelitian menjadi lebih transparan dan dapat dipertanggungjawabkan. Kelebihan dari RapidMiner tersedianya berbagai ekstensi tambahan yang mendukung analisis big data, integrasi dengan basis data relasional, serta kemampuan penerapan metode visualisasi. Namun, kelemahannya adalah keterbatasan fitur pada versi gratis, terutama untuk pemrosesan dataset dalam jumlah besar. Meski begitu, RapidMiner tetap menjadi salah satu pilihan utama bagi peneliti maupun praktisi data mengolah data secara efektif.

2.2.7 Media Sosial

Sarana atau alat yang digunakan untuk menyampaikan pesan, informasi, atau gagasan dari satu pihak sehingga terjadi proses komunikasi. kata *media* berasal dari bahasa Latin yang berarti “perantara” atau “penghubung”. Dalam konteks komunikasi, media berfungsi sebagai jembatan yang memudahkan penyampaian pesan agar dapat diterima dan dipahami. Media dapat dibedakan menjadi beberapa jenis, antara lain media cetak (seperti surat kabar, majalah, dan buku), media elektronik (seperti televisi dan radio), serta media digital (seperti internet, media sosial, dan aplikasi berbasis teknologi). Perkembangan teknologi telah mendorong penggunaan media tradisional ke media digital yang kini lebih interaktif, cepat, dan menjangkau pengguna dalam skala global. Media Sosial dipandang sebagai sarana komunikasi modern yang memungkinkan penggunaannya berbagi informasi dalam bentuk teks, gambar, maupun video.

2.2.8 Pengembangan Sistem



Suatu proses terstruktur yang bertujuan untuk merancang, membangun, mengimplementasikan, serta memelihara sebuah sistem agar dapat memenuhi kebutuhan pengguna maupun organisasi. Sistem dalam konteks ini dapat berupa sistem informasi, perangkat lunak, maupun kombinasi dari komponen teknologi, manusia, dan prosedur yang bekerja bersama untuk mencapai tujuan tertentu. Proses pengembangan sistem umumnya mencakup beberapa tahapan, seperti analisis kebutuhan, perancangan (*design*), implementasi, pengujian (*testing*), hingga pemeliharaan (*maintenance*). Metodologi yang sering digunakan dalam pengembangan sistem antara lain *System Development Life Cycle (SDLC)*, *Agile Development*, dan *Rapid Application Development (RAD)*. Tujuan utama dari pengembangan sistem adalah untuk meningkatkan efektivitas, efisiensi, dan kualitas layanan dengan memanfaatkan teknologi yang tepat. Dalam penelitian ini, pengembangan sistem dilakukan untuk merancang sebuah sistem

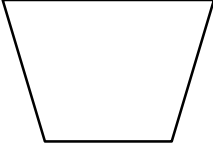
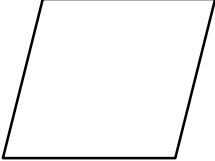

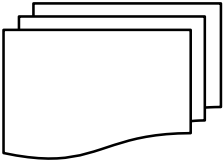

informasi edukatif berbasis analisis sentimen, yang bertujuan memberikan panduan kepada orang tua dan pendidik dalam mendampingi anak usia dini menggunakan media sosial Instagram secara sehat dan bertanggung jawab.

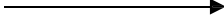
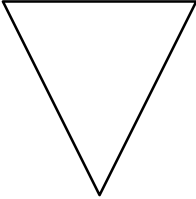
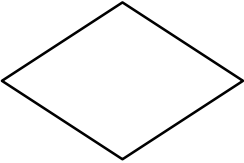
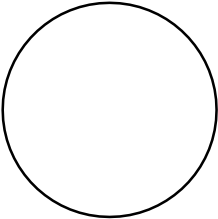
2.2.9 Flowchart

Representasi grafis dari suatu proses atau sistem yang menggambarkan urutan langkah-langkah atau tindakan yang harus diambil untuk mencapai suatu tujuan. Flowchart menggunakan simbol-simbol standar untuk mewakili berbagai jenis tindakan atau langkah, dan garis atau panah untuk menunjukkan urutan dan hubungan antar langkah-langkah tersebut. Flowchart sangat berguna dalam memvisualisasikan proses, membuatnya lebih mudah untuk dipahami, dianalisis, dan dikomunikasikan.[8]

tabel 2.5 Flowchart

Simbol	Nama	Keterangan
	Terminator	Simbol yang digunakan untuk menggambarkan awal atau akhir suatu proses dalam sistem.
	Proses	Simbol yang digunakan untuk menggambarkan proses komputerisasi dari suatu sistem

	Proses Manual	Simbol yang digunakan untuk menggambarkan proses yang dikerjakan manual di dalam sistem.
	Catatan	Simbol yang digunakan untuk menggambarkan catatan data yang direkap sebelumnya didalam dokumen atau formulir.
	Dokumen	Simbol ini menggambarkan sebuah dokumen input atau output yang terkait dalam suatu sistem.
	Multi Dokumen	Simbol ini menunjukkan beberapa dokumen lanjutan
	Penyimpanan	Simbol ini digunakan untuk menggambarkan media penyimpanan yang digunakan dalam suatu sistem

	Garis Alir	Simbol ini adalah symbol yang digunakan untuk menjelaskan arah atau jalanya suatu proses dalam sistem.
	Arsip	Simbol ini menunjukkan dokumen yang diarsipkan.
	Kondisi	Simbol ini digunakan untuk menunjukkan proses pengkondisian dari suatu proses kemudian keputusannya berupa Ya atau Tidak.
	Penghubung	Simbol ini digunakan untuk penghubung ke halaman yang masih sama atau ke halaman lain.

BAB III

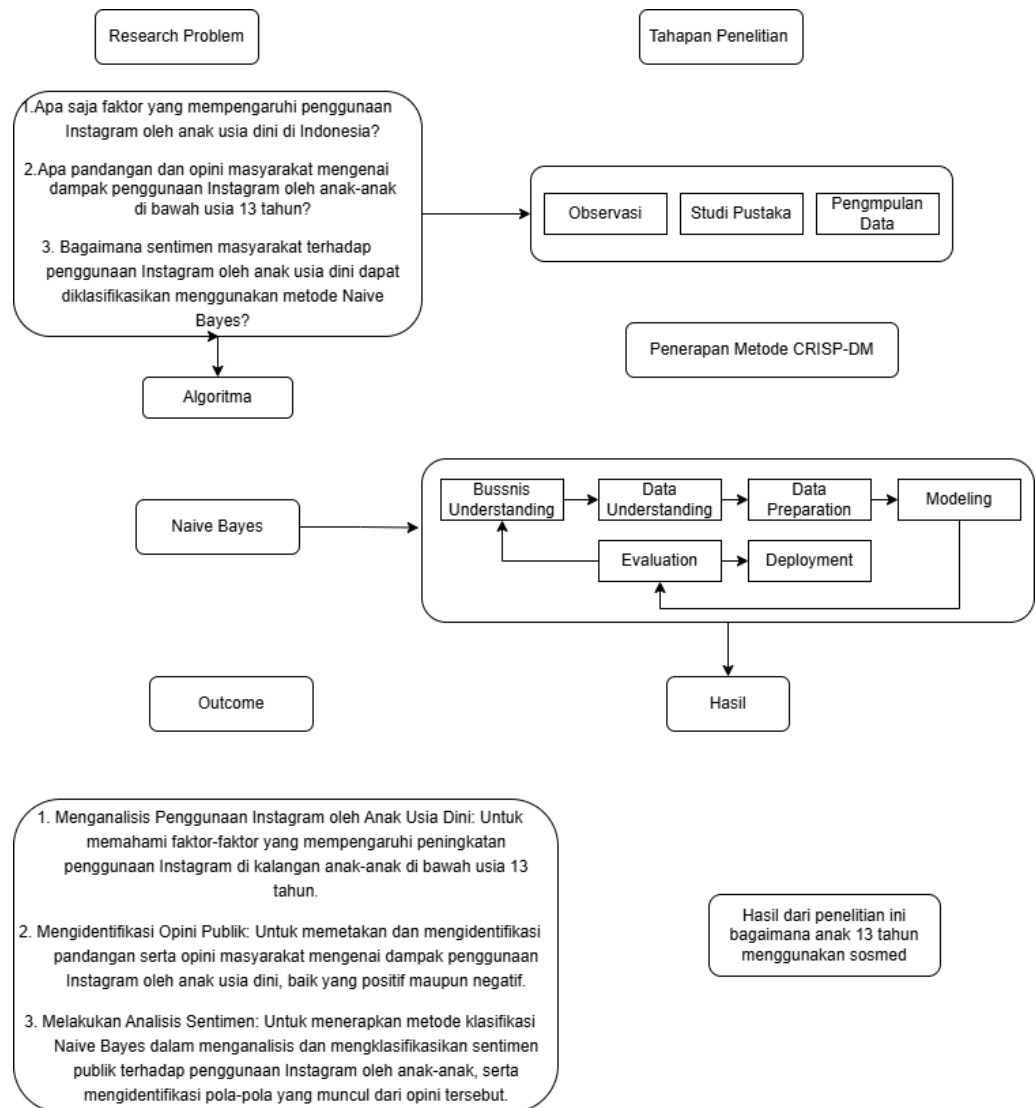
METODE PENELITIAN

3.1 Data Penelitian

Data penelitian ini dikumpulkan secara khusus dari platform Instagram yang membahas tentang penggunaan Instagram oleh anak usia dini. Instagram dipilih sebagai sumber data utama karena platform ini merupakan salah satu media sosial yang paling populer dan banyak digunakan oleh orang tua untuk berbagi konten tentang anak-anak mereka. Data yang diambil dari komentar tentang konten anak usia dini diperoleh berjumlah 435 komentar.[17]

3.2 Penerapan Metode

Pada tahapan penelitian ini akan dilakukan pengumpulan data awal maka proses selanjutnya akan dilakukan penarikan data, kemudian setelah itu dilakukan *Cleansing* Data, lalu pada tahap selanjutnya dilakukan Pemilihan Atribut, tahap berikutnya melakukan penerapan metode CRISP-DM yang pada proses penelitian ini mengacu pada enam tahap CRISP-DM.[13] Berikut tahapan metodologi penelitian yang dapat dilihat pada Gambar 3.1



gambar 3.1 Tahapan Penerapan Metodologi

Penerapan metode dalam penelitian ini dilakukan melalui beberapa tahap yang sistematis, dimulai dari pengumpulan data hingga analisis sentimen menggunakan algoritma *Naive Bayes*. Berikut adalah penjelasan tahap demi tahap dari metode yang digunakan:



gambar 3.2 Kerangka Penelitian

Dalam gambar diatas bisa dijelaskan penelitian ini meliputi 4 proses yang di lakukan yaitu :

- a. *GET Data Instagram*
Objek pada deskripsi data penelitian ini adalah pandangan pengguna Instagram terhadap anak kecil. Konten yang berisi anak kecil
- b. *Cleasing Data*
Hashtag yang belum diolah melalui beberapa proses, yaitu: cleaning, tokenisasi, transform cases, stopword, dan filtering.
- c. *Labeling Data*
Data yang sudah melalui proses cleansing kemudian melakukan pelabelan secara manual. Pelabelan secara manual dilakukan pada seluruh Hashtag.
- d. *Processing Data*
Pada penelitian ini akan dilakukan pengujian menggunakan confusion matrix untuk melihat hasil pengujian data yang

diperoleh dari tahapan modelling dengan menggunakan algoritma *Naïve Bayes* untuk menilai akurasi dari algoritma tersebut.

3.3 Rancangan Pengujian

3.3.1 Business Understanding (Pemahaman Bisnis)

Berfokus pada pemahaman tujuan penelitian terkait penggunaan Instagram oleh anak usia dini. Masalah utama yang diidentifikasi adalah kebutuhan untuk mengklasifikasikan konten Instagram yang aman dan tidak aman bagi anak-anak berusia 3-6 tahun. Tujuan bisnis meliputi pengembangan sistem filtering otomatis yang dapat melindungi anak-anak dari konten tidak pantas, membantu orang tua dalam mengawasi aktivitas digital anak, dan menciptakan lingkungan media sosial yang lebih aman. Peran utama dalam proyek ini adalah orang tua, pendidik anak usia dini, pengembang aplikasi parental control, dan platform media sosial itu sendiri yang ingin meningkatkan keamanan pengguna muda mereka.[15]

3.3.2 Data Understanding (Pemahaman Data)

Pada tahap ini dilakukan eksplorasi dan analisis terhadap data Instagram yang akan digunakan untuk pelatihan model *Naïve Bayes*. Data yang dikumpulkan mencakup teks caption, hashtag, komentar, metadata gambar, dan informasi profil pengguna dari postingan Instagram. Karakteristik data menunjukkan variasi bahasa (Indonesia dan Inggris), emoji, *cyber slang*, dan berbagai jenis konten multimedia. [18]

3.3.3 Data Preparation (Persiapan Data)

Proses dimulai dengan pembersihan teks yang meliputi penghapusan URL, *hashtag* berlebihan, dan karakter khusus yang tidak relevan. Normalisasi teks dilakukan melalui konversi ke huruf kecil, penghapusan stopwords bahasa Indonesia dan Inggris.

Labeling data dilakukan secara manual untuk memastikan akurasi kategori "aman" dan "tidak aman" berdasarkan standar konten yang sesuai untuk anak usia dini.[14]

3.3.4 Modeling (Pemodelan)

Tahap pemodelan akan menggunakan metode *classification* dengan algoritma *Naïve Bayes*. Dalam penerapan metode *classification* akan diterapkan konten anak kecil. Alat pemodelan yang digunakan adalah Rapidminer.[19]

3.3.5 Evaluation (Evaluasi)

Pada tahap ini akan dilakukan analisa atau pengukuran ketepatan terhadap pemodelan yang telah dilakukan. Evaluasi ditujukan untuk mengetahui pemodelan yang dilakukan apakah sudah tepat dan sesuai diterapkan pada kasus penelitian ini serta sudah sesuai rencana awal penelitian. Selanjutnya dari hasil evaluasi tersebut adalah menentukan langkah berikutnya apakah bisa dilanjutkan atau diulang dari awal karena tidak sesuai dengan rencana awal penelitian.[20]

3.3.6 Deployment

Tahap selanjutnya yaitu tahap penyebaran hasil dari penelitian yang sudah dilakukan dijadikan sebagai laporan atau presentasi dari pengetahuan yang telah didapat berdasarkan pemodelan dan evaluasi pada proses data mining.

BAB IV

HASIL YANG DIHARAPKAN

4.1 Alat Penelitian

Dalam penelitian, membutuhkan alat untuk mendukung berjalannya penelitian. Alat penelitian yang digunakan yaitu berupa perangkat lunak (*software*) dan perangkat keras (*hardware*).

4.1.1 Perangkat lunak (*software*)

Berikut alat perangkat lunak (*software*), versi, dan fungsinya dapat dilihat pada tabel 4.1 :

tabel 4.1 Spesifikasi Software

No	Jenis	Versi	Fungsi
1	Microsoft Excel	2010	Untuk memfilter data dan menyimpan data
2	Rapidminer	10.2	Untuk melakukan pengolahan data
3	Mendeley	1.19.8	Memudahkan proses penulisan karya ilmiah, mulai dari mencari referensi, hingga membuat kutipan dan daftar pustaka secara otomatis.

4.1.2 Perangkat Keras (*hardware*)

Berikut alat perangkat keras (*hardware*), jenis dan spesifikasinya dapat dilihat pada tabel 4.2 :

tabel 4.2 Spesifikasi Hardware

No	Jenis	Spesifikasi
1	<i>Processor</i>	Inter(R) Celeron(R)N420 CPU 1.010Ghz (2CPUs)
2	RAM	4GB
3	<i>System Type</i>	Windows
4	<i>Operating System</i>	Windows 11 Home

4.2 Implementasi Dan Pengujian

Implementasi dan pengujian menggunakan metode Naïve Bayes dilakukan berdasarkan data dari komentar pada konten anak usia dini 435 komentar *positif*, *negative* dan *netral*.

a. Pengumpulan Data

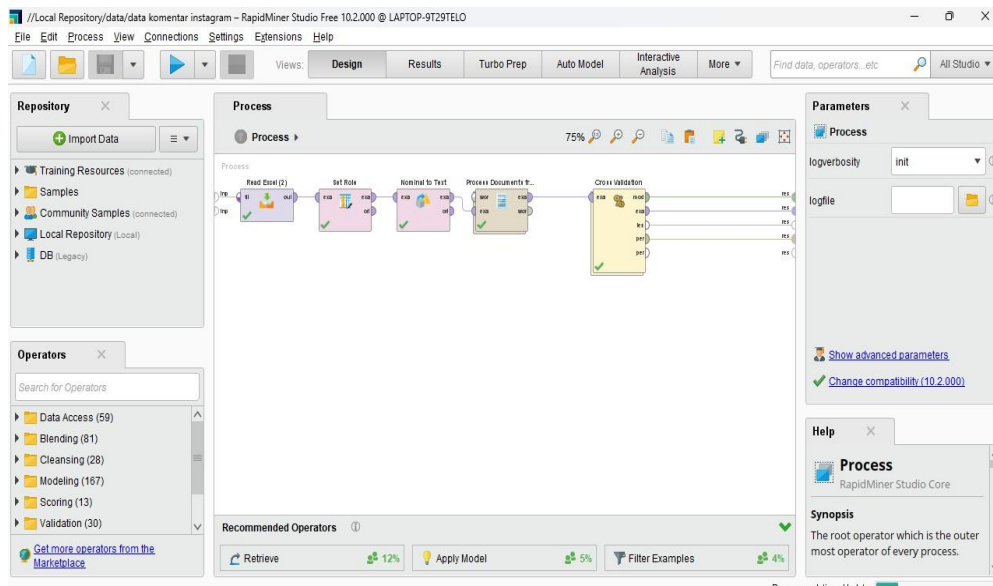
Untuk pengambilan data menggunakan excel Jumlah data yang dianalisis dalam penelitian ini Adalah 435 terdiri dari positif, negative dan netral digambarkan pada gambar 4.1

	A	B	C	D	E
1	nomor	Bahasa Baku			
2	1	menggunakan jaket kita ke rumah rumah sakit jiwa	negatif		
3	2	kerusakannya itu loh sangat tidak ada	negatif		
4	3	biarin saja nanti juga besar malu sendiri	negatif		
5	4	bertato saja bangga palingan juga dari permen karet	negatif		
6	5	kamu kalau mau garuk idung bilang tidak usah di rekam video	negatif		
7	6	sangat tidak nyambung banget	negatif		
8	7	wah sangat keren dan sangat cantik	positif		
9	8	auranya itu loh sangat tidak ada	negatif		
10	9	seperti orang gila bikin kayak seperti itu	negatif		
11	10	sungguh tidak menarik	negatif		
12	11	mengapa dengan dia	negatif		
13	12	untuk apa dia berperilaku seperti itu	negatif		
14	13	saya sangat merinding asli	negatif		
15	14	sungguh menjijikan membuat seperti itu	negatif		
16	15	suara dan nada terpisah	negatif		
17	16	yang tertawa bukan hanya langit	positif		
18	17	nada sesuai kepercayaan masing-masing	positif		
19	18	besok akan menjadi meme ini	negatif		
20	19	saya tidak mau kembali kesini lagi	negatif		
21	20	apa malam malam bikin heboh	negatif		
22	21	ini ceritanya membuat apa	negatif		
23	22	aduh sangat takut kepada adik ini	negatif		

gambar 4.1 jumlah data yang di dapat



b. Cleansing Data

Pada tahapan ini peneliti akan melakukan proses Cleansing dengan menggunakan rapidminer sebelum melakukan pemberian label pada sentimen, agar dataset pada file excel terbaru terhindar dari duplikasi data dan tanda yang tidak diperlukan seperti yang digambarkan pada gambar 4.2.



gambar 4.2 Model *Proses Cleansing*

Setelah selesai menjalankan proses *Cleansing* dengan menggunakan rapidminer data yang di dapat seperti pada gambar 4.3

Open in  Turbo Prep  Auto Model Filter (434 / 434 examples): all

Row No.	hasil	text	abang	abi	abisan	adakah	adalah	adik
1	negatif	menggunaka...	0	0	0	0	0	0
2	negatif	kerusakanny...	0	0	0	0	0	0
3	negatif	biarin saja na...	0	0	0	0	0	0
4	negatif	bertato saja b...	0	0	0	0	0	0
5	negatif	kamu kalau g...	0	0	0	0	0	0
6	negatif	sangat tidak ...	0	0	0	0	0	0
7	positif	sangat keren ...	0	0	0	0	0	0
8	negatif	auranya sang...	0	0	0	0	0	0
9	negatif	seperti orang...	0	0	0	0	0	0
10	negatif	sungguh tida...	0	0	0	0	0	0
11	negatif	mengapa de...	0	0	0	0	0	0
12	negatif	untuk berpel...	0	0	0	0	0	0

gambar 4.3 Hasil *Cleansing Data*

c. Labeling

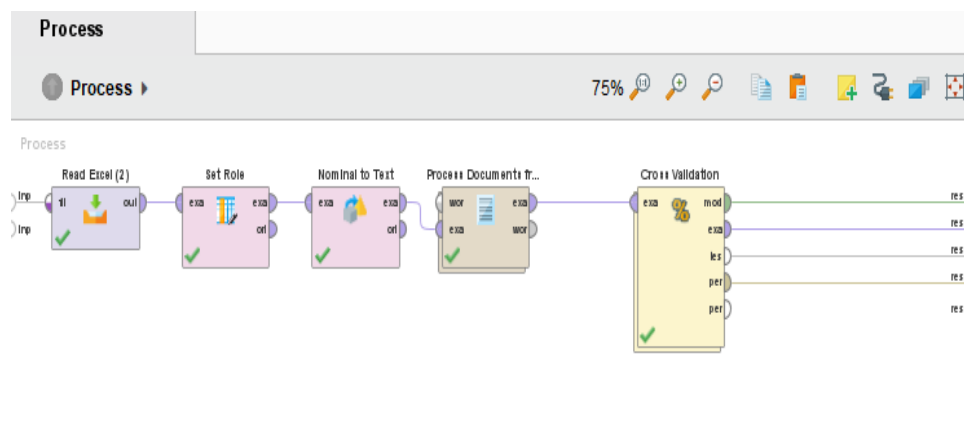
Kemudian tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentiment berlangsung. Berikut data yang sudah di beri label seperti pada gambar 4.4.

	B	C	D	E	F
1	Bahasa Baku				
2	menggunakan jaket kita ke rumah rumah sakit jiwa	negatif			
3	kerusakannya itu loh sangat tidak ada	negatif			
4	biarin saja nanti juga besar malu sendiri	negatif			
5	bertato saja bangga palingan juga dari permen karet	negatif			
6	kamu kalau mau garuk idung bilang tidak usah di rekam video	negatif			
7	sangat tidak nyambung banget	negatif			
8	wah sangat keren dan sangat cantik	positif			
9	auranya itu loh sangat tidak ada	negatif			
10	seperti orang gila bikin kayak seperti itu	negatif			
11	sungguh tidak menarik	negatif			
12	mengapa dengan dia	negatif			
13	untuk apa dia berperilaku seperti itu	negatif			
14	saya sangat merinding asli	negatif			
15	sungguh menjijikan membuat seperti itu	negatif			
16	suara dan nada terpisah	negatif			
17	yang tertawa bukan hanya langit	positif			
18	nada sesuai kepercayaan masing-masing	positif			
19	besok alcan menjadi meme ini	negatif			
20	saya tidak mau kembali kesini lagi	negatif			
21	apa malam malam bikin heboh	negatif			
22	ini ceritanya membuat apa	negatif			
23	aduh sangat takut kepada adik ini	negatif			
24	nanti jika dewasa jangan seperti ini ya dek	negatif			
25	dik minuman itu membuat menjadi beban lho	negatif			
26	sudah minum jamu buyung upi saja de tidak usah minum itu	positif			

gambar 4.4 data yang sudah dilabeli

d. Tokenizing

Pada proses ini adalah untuk mengubah teks dari kalimat-kalimat yang kompleks menjadi urutan token atau unit-unit yang lebih sederhana. Tujuannya untuk mempersiapkan data teks agar dapat diproses oleh algoritma analisis sentimen dengan lebih efektif dan akurat. Untuk proses *tokenizing*, Hasil *Tokenizing Naïve bayes* dapat dilihat sebagaimana yang ada pada gambar 4.5, gambar 4.6 dibawah ini.



gambar 4.5 Model *Tokenizing* Data

PerformanceVector (Performance (2))

Table View Plot View

accuracy: 47.93% +/- 5.68% (micro average: 47.93%)

	true negatif	true positif	true netral	true negatif	class prec
pred. negatif	52	37	26	2	44.44%
pred. positif	48	97	20	0	58.79%
pred. netral	38	27	43	6	37.72%
pred. negatif	7	4	11	16	42.11%
class recall	35.86%	58.79%	43.00%	66.67%	

gambar 4.6 Hasil *Tokenizing* Naïve bayes

4.3 Results

Berikut dideskripsikan olah aplikasi *rapidminer* seperti yang ada pada gambar 4.7

PerformanceVector

PerformanceVector:

accuracy: 47.93% +/- 5.68% (micro average: 47.93%)

ConfusionMatrix:

```

True:  negatif          positif netral  negatif
negatif :      52      37      26      2
positif:      48      97      20      0
netral: 38      27      43      6
negatif:      7       4      11     16

```

gambar 4.7 Deskripsi Akurasi *Naïve bayes*

SimpleDistribution

Distribution model for label attribute hasil

Class negatif (0.334)
1026 distributions

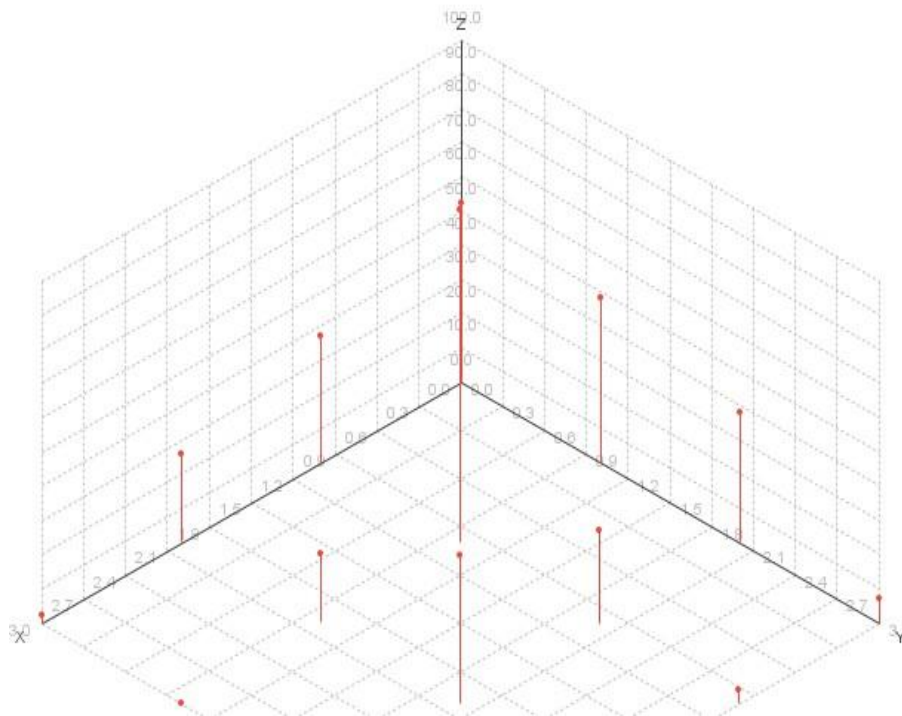
Class positif (0.380)
1026 distributions

Class netral (0.230)
1026 distributions

Class negatif (0.055)
1026 distributions

gambar 4.8 Deskripsi *Simple Distribution*

Dari hasil tersebut tergambar *plotview* untuk algoritma *Naïve bayes* seperti yang digambarkan pada gambar 4.9 dibawah ini



gambar 4.9 *Plot View Naïve bayes*

Table View Plot View

accuracy: 47.93% +/- 5.68% (micro average: 47.93%)

	true negatif	true positif	true netral	true negatif	class pre
pred. negatif	52	37	26	2	44.44%
pred. positif	48	97	20	0	58.79%
pred. netral	38	27	43	6	37.72%
pred. negatif	7	4	11	16	42.11%
class recall	35.86%	58.79%	43.00%	66.67%	

gambar 4.10 *Table View*

Result History PerformanceVector (Performance (2))

Result History: Display a list of results of this session's processes.

data komentar instagram (3 results, Process results)
Completed: Aug 28, 2025 1:14:34 PM (execution time: 13 s)

Distribution Model (Naive Bayes)
Result not stored in repository.

Distribution model for label attribute

Class negatif (0.334)
1026 distributions

Class positif (0.380)
1026 distributions

Class netral (0.230)
1026 distributions

Class negatif (0.055)

ExampleSet (Process Documents from Data)
Result not stored in repository.

Data Table
Number of examples = 434
At least 250 attributes:

Role	Name	Type	Range	Missings	Con
=				no	
-	abang	real	[?...?]; mean	missing values	-
=				no	
-	abi	real	[?...?]; mean	missing values	-
=				no	
-	abi	real	[?...?]; mean	missing values	-

Performance Vector (Performance (2))
Result not stored in repository.

PerformanceVector:
accuracy: 47.93% +/- 5.68% (micro average)

ConfusionMatrix:

True:	negatif	positif	netral	r
negatif :	52	37	26	2
positif:	48	97	20	0
netral:	38	27	43	6
negatif:	7	4	11	1

gambar 4.11 Result history

Attribute	Parameter	negatif	positif	netral	negatif
abang	mean	0	0.003	0	0
abang	standard deviation	0.001	0.045	0.001	0.001
abi	mean	0	0	0.005	0
abi	standard deviation	0.001	0.001	0.048	0.001
abisan	mean	0	0	0.005	0
abisan	standard deviation	0.001	0.001	0.048	0.001
adakah	mean	0.005	0	0	0
adakah	standard deviation	0.041	0.001	0.001	0.001
adalah	mean	0.005	0.011	0	0
adalah	standard deviation	0.034	0.062	0.001	0.001
adik	mean	0.010	0.007	0.017	0
adik	standard deviation	0.057	0.049	0.083	0.001
adiknva	mean	0.008	0.007	0.007	0

gambar 4.12 Distribution table

Name	Type	Missing	Filter (1,028 / 1,028 attributes):	Search for Attributes	
Label hasil	Polynomial	0	Least negatif (24)	Most positif (165)	Value posi
Text text	Text	0	Least yang ter [...] ngit (1)	Most bagu sek [...] main (3)	Value bag
abang	Real	0	Min 0	Max 0.573	Average 0.00
abi	Real	0	Min 0	Max 0.481	Average 0.00
abisan	Real	0	Min 0	Max 0.481	Average 0.00
adakah	Real	0	Min 0	Max 0.410	Average 0.00

gambar 4.13 *Statistic*

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan analisis sentimen yang dilakukan terhadap media sosial, khususnya penggunaan Instagram oleh anak usia dini, dapat ditarik beberapa kesimpulan sebagai berikut:

- a. Terdapat berbagai pandangan dan opini di masyarakat mengenai dampak penggunaan Instagram oleh anak usia dini. Opini ini terbagi menjadi tiga kategori utama, yaitu positif, negative.
- b. Penggunaan Instagram oleh anak usia dini dipengaruhi oleh berbagai faktor, termasuk akses mudah ke teknologi, pengaruh teman sebaya, serta kurangnya pengawasan dan pemahaman dari orang tua dan pendidik.
- c. Dampak Penggunaan Instagram: Penggunaan Instagram oleh anak-anak dapat menimbulkan dampak positif dan negatif. Dampak negatif meliputi paparan konten yang tidak sesuai, risiko cyberbullying, serta gangguan dalam perkembangan sosial dan psikologis.
- d. Metode Naïve Bayes terbukti efektif dalam menganalisis sentimen dan mengklasifikasikan opini publik terkait penggunaan Instagram oleh anak usia dini. Algoritma ini mampu memberikan wawasan berharga bagi pengembangan sistem informasi edukatif.
- e. Pengembangan sistem informasi edukatif berbasis analisis sentimen dapat memberikan informasi komprehensif dan panduan bagi orang tua dan pendidik dalam mendampingi anak-anak menggunakan media sosial secara sehat dan bertanggung jawab.

5.2 Saran

Berdasarkan kesimpulan di atas, berikut adalah beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya dan implementasi praktis:

- a. Prioritaskan pengembangan sistem informasi edukatif yang mudah diakses dan dipahami oleh orang tua dan pendidik.
- b. Sertakan fitur-fitur seperti panduan pengaturan privasi, edukasi konten sesuai usia, serta strategi untuk meminimalkan dampak negatif penggunaan berlebihan.
- c. Orang tua dan pendidik perlu meningkatkan pengawasan dan pendampingan terhadap penggunaan media sosial oleh anak-anak.
- d. Edukasi mengenai risiko dan manfaat media sosial perlu diberikan secara berkala kepada anak-anak.

Daftar Pustaka

- [1] D. R. Aditya, E. Supriyati, and T. Listyorini, “Analisis Sentimen Pengguna Twitter Terhadap Rokok Elektrik (Vape) Di Indonesia Menggunakan Metode Naïve Bayes,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.,* vol. 7, no. 1, pp. 43–50, 2022, doi: 10.29100/jipi.v7i1.2145.
- [2] M. Ula and S. Fachrurrazi, “Analisis Sentimen Cyberbullying pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Naïve Bayes Classifier,” *TECHSI - J. Tek. Inform.,* vol. 14, no. 2, p. 91, 2023, doi: 10.29103/techsi.v14i2.12103.
- [3] N. D. Pratidina and J. Mitha, “Dampak Penggunaan Media Sosial terhadap Interaksi Sosial Masyarakat: Studi Literature,” *J. Ilm. Univ. Batanghari Jambi,* vol. 23, no. 1, p. 810, 2023, doi: 10.33087/jiubj.v23i1.3083.
- [4] N. A. Susanti, M. Walid, and H. Hoiriyah, “Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial Menggunakan Naive Bayes Classifier,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 2, pp. 538–543, 2022, doi: 10.36040/jati.v6i2.5174.
- [5] F. Zamzami, R. Hidayat, and R. Fathonah, “Penerapan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Komentar Twitter Proyek Pembagunan Ikn,” *Fakt. Exacta,* vol. 17, no. 1, pp. 47–57, 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [6] J. Marzuki and P. Sibarani, “Implementation of RFM Analysis to Enhance Sales Patterns of Food and Beverages at Bonjour Café and Resto Using the Apriori Algorithm,” vol. 4, no. December, pp. 1261–1279, 2024.
- [7] R. Sanjaya, Edi Tohidi, and E. Wahyudi, “Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” vol. 8, no. 1, pp. 507–514, 2024.
- [8] A. Anjani, A. Chamid, and A. Murti, “Analisis Sentimen Kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” *JTINFO J. Tek. Inform.,* vol. 1, no. 2, pp. 1–8, 2022, [Online]. Available: <https://journal.unisnu.ac.id/JTINFO/article/view/259>

- [9] A. C. Devi and U. M. Fadli, “Analisis Keputusan Pembelian Produk E-Commerce Shopee Di Kalangan Mahasiswa Universitas Buana Perjuangan Karawang,” *J. Student Res.*, vol. 1, no. 5, pp. 113–123, 2023.
- [10] T. A. Zuraiyah, M. M. Mulyati, and G. H. F. Harahap, “Perbandingan Metode Naïve Bayes, Support Vector Machine Dan Recurrent Neural Network Pada Analisis Sentimen Ulasan Produk E-Commerce,” *Multitek Indones.*, vol. 17, no. 1, pp. 27–43, 2023, doi: 10.24269/mtkind.v17i1.7092.
- [11] E. Laia and M. Yamin, “Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce,” *Media Online*, vol. 4, no. 1, pp. 305–316, 2023, doi: 10.30865/klik.v4i1.1186.
- [12] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [13] S. Adelia, E. Milanda, J. Santari, D. T. Kesuma, E. Silvia, and F. Kurniawan, “Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes,” *J. Inf. Technol. Ampera*, vol. 4, no. 3, pp. 254–264, 2023, [Online]. Available: <https://journal-computing.org/index.php/journal-ita/article/view/430>
- [14] K. Cyberbullying and D. I. Instagram, “Komparasi metode algoritma klasifikasi pada analisis sentimen komentar cyberbullying di instagram,” vol. 7, pp. 466–472, 2024, doi: 10.37600/tekinkom.v7i1.1392.
- [15] O. P. Zusrotun, A. C. Murti, and R. Fiati, “Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes,” *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.
- [16] S. M. Harahap and R. Kurniawan, “Analisis Sentimen Komentar Youtube Terhadap Food Vlogger Dengan Menggunakan Metode Naïve Bayes,” *MEANS (Media Inf. Anal. dan Sist.*, vol. 9, no. 1, pp. 87–96, 2024, doi: 10.54367/means.v9i1.3912.

- [17] R. P. M. Rosidi and K. Setiawan, “Implementasi Algoritma Naïve Bayes Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada Kantin,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 120–126, 2024, doi: 10.35870/jimik.v5i1.407.
- [18] W. Wahyuni, “Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes,” *J. Inform. Ekon. Bisnis*, vol. 4, pp. 148–153, 2022, doi: 10.37034/infeb.v4i4.162.
- [19] P. K. D. I. E-commerce, “BIG DATA ANALYTICS DAN MACHINE LEARNING UNTUK MEMPREDIKSI,” vol. 8, no. 1, pp. 159–167, 2025.
- [20] S. Heristian, M. Napiah, and W. Erawati, “Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek,” vol. 5, no. 1, pp. 35–41, 2025.



Surat Penunjukan Pembimbing Skripsi

Berdasarkan Keputusan Ketua Sekolah Tinggi Ilmu Komputer Cipta Karya

Informatika No. : 311/SK/K/STI KOMCKI/IV/2025

Dengan ini menunjuk Bapak/Ibu sebagai pembimbing SKRIPSI atas

Nama Mahasiswa : Salsabila Putri Wibowo

NIM : 21120110025

Program Studi : Sistem Informasi

Judul SKRIPSI, Analisis sentiment media sosial terhadap penggunaan Instagram oleh anak usia dini menggunakan metode *Naïve Bayes* sebagai dasar pengembangan sistem informasi edukatif

Jakarta, 10 mei 2025

Menyetujui,

Ir Sugiono M, Kom
NIDN : 0327116002

Mahasiswa,

Salsabila Putri W

Mengetahui,

**Ketua Program Studi
Sistem Informasi**

Veri Arinal, M.Kom
NIDN : 0328047101

Kepala Bagian Keuangan

Muchammad Zaeny, SE, M. Pd
NIP : 202105



Permohonan Penulisan Bimbingan Skripsi

Nama Mahasiswa : Salsabila Putri Wibowo

NIM : 21120110025

Program Studi : Sistem Informasi

Dengan ini memohon untuk dapat menulis SKRIPSI, dengan judul/topik adalah Analisis sentiment media sosial terhadap penggunaan Instagram oleh anak usia dini menggunakan metode *Naïve Bayes* sebagai dasar pengembangan sistem informasi edukatif

Jakarta, 10 mei 2025

Menyetujui,
Ketua Program Studi

Veri Arinal, M.Kom
NIDN : 0328047101

Mahasiswa,

Salsabila Putri W

(Form-3)

**KARTU BIMBINGAN DAN KONSULTASI SKRIPSI**

Nama Mahasiswa : Salsabila Putri Wibowo
 NIM : 21120110025
 Program Studi : Sistem Informasi
 Nama Pembimbing : Ir. Sugiono, M.Kom


No.	Hari/Tanggal	Uraian Kegiatan/ Saran-Saran	Tanggal Kembali	Paraf Pembimbing
1	19 Mei 2025	Bimbingan Bab 1		
2	28 Mei 2025	Bimbingan Bab 2		
3	14 Juni 2025	Bimbingan Bab 3		
4	30 Juni 2025	Bimbingan Bab 3		
5	9 Agustus 2025	Bimbingan Bab 5 & 4		
6	12 Agustus 2025	Bimbingan bab 5 & ace skripsi		

Proses Bimbingan :

1. Dimulai Tanggal : 19 Mei 2025

2. Berakhir Tanggal : 12 Agustus 2025

Ttd Pembimbing,


 (Ir. Sugiono, M.Kom)



STIKOM
CIPTA KARYA INFORMATIKA

TANDA PERSETUJUAN

Nama : Salsabila Putri W
Nim : 21120110025
Program Studi/Program : Sistem Informasi/S-1
Judul SKRIPSI : Analisis Sentimen Media Sosial Terhadap Penggunaan Instagram Oleh Anak Usia Dini Pengembangan Sistem Informasi Edukatif Berbasis Naïve bayes

Jakarta, 14 Agustus 2025

Menyetujui,
Pembimbing

(Ir Sugiono M, Kom.
NIDN : 0327116002

Mengetahui,
Ketua Program Studi
Sistem Informasi

(Veri Arinal, S.Kom, M.Kom.)
NIDN : 0328047101

Wakil Ketua Bid. Akademik

Yuma Akbar, S.Kom, M. Kom
NIDN : 0319038304

revisi salsabila

ORIGINALITY REPORT

29% SIMILARITY INDEX	24% INTERNET SOURCES	19% PUBLICATIONS	10% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

PRIMARY SOURCES

1	Aji Dwi Prasetyo, Mesra Betty Yel. "Analisa Sentimen Menggunakan Algoritma C4.5 Dan Naïve Bayes Dengan Topik Artificial Intelligence Pada Media Sosial Twitter (X)", INTECOMS: Journal of Information Technology and Computer Science, 2024 Publication	2%
2	ejournal.sisfokomtek.org Internet Source	2%
3	eprints.unisbank.ac.id Internet Source	2%
4	repository.bsi.ac.id Internet Source	1%
5	journal.amikindonesia.ac.id Internet Source	1%
6	ejournal.itn.ac.id Internet Source	1%
7	ejournal.ust.ac.id Internet Source	1%
8	Suci Sugih Hartinah, Sugiyono. "Pemodelan Data Mining Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus: Kedai Ngodeng & Smoothies)", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2024	1%

RIWAYAT HIDUP

A. Data Pribadi

- Nama : Salsabila Putri Wibowo
- Tempat, Tanggal Lahir: Jakarta, 27 juni 2003
- Jenis Kelamin : Perempuan
- Alamat : Jl. Pkp Raya RT 003/009 ,Kelapa Dua Wetan, Ciracas
- No. Telepon : 088295652433
- Email : slsblpw95@gmail.com

B. Riwayat Pendidikan

1. SDN 03 CIBUBUR [2009-2015]
2. MTS.N 22 Munjul [2015-2018]
3. SMK Paskita Global [2018-2021]
4. MAHASISWA AKTIF SEMESTER VIII DI STIKOMCKI [2021- 2025]

C. Pengalaman Kerja

1. Part time di S2 Banquet
2. Part Time waitress di hotel Whizprime
3. Barista di Pinarak Coffe
4. Barista di Mr.Signature

D. Keterampilan

- Public Speaking & Teamwork

E. Prestasi

- Mendapat Mendali Emas Dalam Lomba Club Taekwondo (2014)

Bandung, 29 Agustus 2025

Hormat saya,



Salsabila Putri W