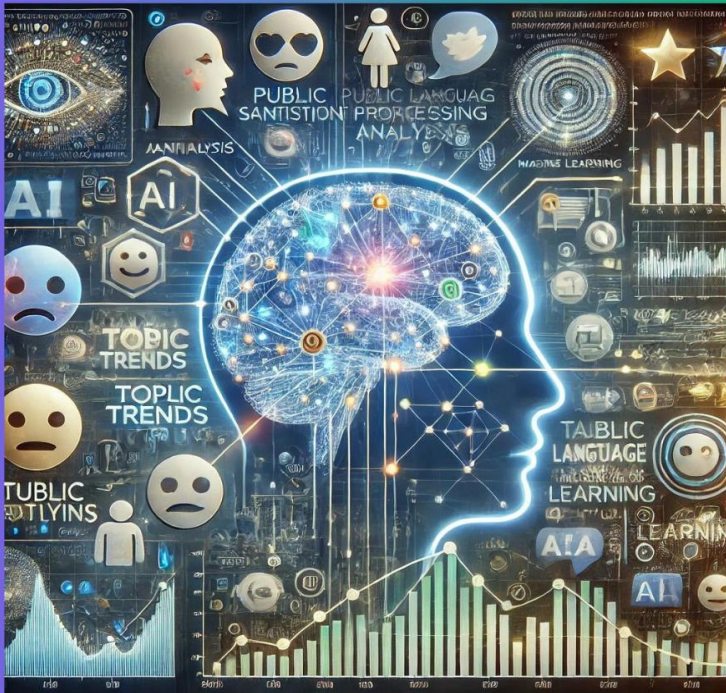


# MONOGRAF

Analisis Sentimen Publik  
dan Tren Topik Pendekatan  
*Natural Language  
Processing* dan *Machine  
Learning* dalam Era Digital



Shabrina Rasyid Munthe, S.Pdi M.S

Sudi Suriadi S.Kom M.Kom

Dr. Basyarul Ulya, SH., MM

Muhammad Rusli, MS

Nurhanna Harahap, M.Hum

Muhammad Fauzi Romadhon Marpaung, S.Pd.I, M.Hum

Raden Sri Ayu Ramadhana, S.Pd .I, M.Pd

Ibnu Rasyid Munthe, ST., M. Kom

*Monograf*  
**Analisis Sentimen Publik dan Tren  
Topik Pendekatan *Natural  
Language Processing* dan  
*Machine Learning* dalam Era  
Digital**

Shabrina Rasyid Munthe, S.Pdi M.S  
Sudi Suriadi S.Kom M.Kom  
Dr. Basyarul Ulya, SH., MM  
Muhammad Rusli, MS  
Nurhanna Harahap, M.Hum  
Muhammad Fauzi Romadhon Marpaung, S.Pd.I, M.Hum  
Raden Sri Ayu Ramadhana, S.Pd .I, M.Pd  
Ibnu Rasyid Munthe, ST., M. Kom



2024

---

**MONOGRAF ANALISIS SENTIMEN PUBLIK DAN TREN TOPIK  
PENDEKATAN NATURAL LANGUAGE PROCESSING DAN  
MACHINE LEARNING DALAM ERA DIGITAL**

---

v + 70 hlm : 14 x 20 cm

---

**ISBN: 978-623-8760-17-6**

**Penulis** : Shabrina Rasyid Munthe, S.Pdi M.S  
Sudi Suriadi S.Kom M.Kom  
Dr. Basyarul Ulya, SH., MM  
Muhammad Rusli, MS  
Nurhanna Harahap, M.Hum  
Muhammad Fauzi Romadhon Marpaung, S.Pd.I, M.Hum  
Raden Sri Ayu Ramadhana, S.Pd .I, M.Pd  
Ibnu Rasyid Munthe, ST., M. Kom

**Layout** : Dr. Fidy Arie Pratama, M.Pd

**Desain Sampul** : Arr Rad Pratama

**Cetakan 1** : September 2024

Copyright © 2024 by Penerbit PT Arr rad Pratama  
All rights reserved

Hak Cipta dilindungi Undang-Undang No 19 Tahun 2002.  
Dilarang memperbanyak atau memindahkan sebagian atau seluruh  
isi buku ini dalam bentuk apapun, baik secara elektris maupun  
mekanis, termasuk memfotocopy, merekam atau dengan sistem  
penyimpanan lainnya, tanpa izin tertulis dari Penulis dan Penerbit.

**Isi di luar tanggung jawab percetakan**

Penerbit PT Arr Rad Pratama  
Anggota IKAPI Nomor 485/JBA/2023  
Gedung Nurul Yaqin Cirebon – Jawa Barat Indonesia 45151  
Cirebon Telp. 085724676697  
e-mail: ptarradpratama@gmail.com  
Web : <https://arradpratama.com/>

## KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas terselesaikannya buku **Monograf Analisis Sentimen Publik dan Tren Topik Pendekatan *Natural Language Processing* dan *Machine Learning* dalam Era Digital**. Buku ini hadir sebagai wujud dari kebutuhan yang semakin meningkat akan pemahaman mendalam terkait analisis sentimen publik dan tren topik di era digital. Seiring dengan perkembangan teknologi dan meluasnya penggunaan media sosial, kemampuan untuk memahami opini dan sentimen masyarakat menjadi sangat penting bagi berbagai sektor, termasuk bisnis, politik, hingga sosial kemasyarakatan.

Buku ini menyajikan pendekatan yang komprehensif dan sistematis dalam menganalisis sentimen dan tren topik melalui teknik-teknik *Natural Language Processing* (NLP) dan *Machine Learning* (ML). Kedua pendekatan ini memungkinkan kita untuk mengolah dan memahami data dalam skala besar secara otomatis, yang sebelumnya memerlukan banyak waktu dan tenaga jika dilakukan secara manual. Kami berharap, buku ini dapat memberikan panduan praktis bagi para peneliti, akademisi, praktisi, dan siapa saja yang tertarik dalam

memahami dinamika opini publik serta bagaimana tren terbentuk dan berkembang di dunia digital.

Dalam buku ini, pembaca akan diajak untuk memahami dasar-dasar teori yang mendasari NLP dan *machine learning*, serta bagaimana kedua teknologi ini diterapkan dalam analisis sentimen publik dan tren topik. Contoh-contoh nyata serta implementasi di berbagai sektor juga disajikan agar dapat memberikan gambaran yang lebih jelas dan aplikatif.

Kami menyadari bahwa ilmu pengetahuan terus berkembang, terutama dalam bidang teknologi digital yang sangat dinamis. Oleh karena itu, kami membuka diri terhadap kritik dan saran untuk penyempurnaan di masa mendatang. Harapan kami, buku ini dapat menjadi kontribusi berarti bagi pengembangan ilmu pengetahuan dan teknologi di Indonesia, serta membantu para pembaca untuk semakin memahami pentingnya analisis data di era informasi yang serba cepat ini.

Akhir kata, semoga buku ini bermanfaat dan dapat memberikan inspirasi bagi para pembaca.

Selamat membaca!

Labuhan Batu, September 2024

Tim Penulis

## DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	viii
BAB I PENDAHULUAN	1
BAB II KONSEP SLANG	6
BAB III MEDIA SOSIAL DAN PENGUNAANNYA	9
BAB IV <i>NATURAL LANGUAGE PROCESSING</i>	22
BAB V ANALISIS SENTIMEN PUBLIK DAN TREN TOPIK PENDEKATAN <i>NATURAL LANGUAGE PROCESSING</i> DAN <i>MACHINE LEARNING</i> DALAM ERA DIGITAL	26
DAFTAR PUSTAKA	59
PROFIL PENULIS	65



# BAB I

## PENDAHULUAN

---

Dalam era digital yang terus berkembang, media sosial telah memainkan peran penting sebagai pusat komunikasi dan pertukaran informasi. Pengguna kerap memanfaatkan bahasa informal, termasuk penggunaan kata slang, sebagai sarana untuk mengekspresikan diri secara lebih fleksibel dan personal. Kata-kata slang memiliki signifikansi yang besar dalam komunikasi modern karena mampu mencerminkan identitas individu, keunikan budaya, serta dinamika sosial dari kelompok tertentu. Dalam interaksi sosial, penggunaan slang dapat memperkuat ikatan antaranggota kelompok, menciptakan rasa kebersamaan, dan memfasilitasi pemahaman yang lebih mendalam di antara mereka. Selain itu, kreativitas dalam penggunaan slang seringkali mencerminkan perubahan dan evolusi bahasa yang bersifat dinamis, menunjukkan bagaimana bahasa terus

berkembang sesuai dengan perkembangan sosial dan budaya di masyarakat.

Media sosial memainkan peran yang sangat signifikan dalam penyebaran kata slang, terutama melalui platform seperti Twitter, Instagram, dan Facebook. Platform-platform ini memungkinkan penyebaran kata slang dengan cepat dan lintas batas, baik geografis maupun demografis. Media sosial tidak hanya berfungsi sebagai tempat munculnya kata-kata slang baru, tetapi juga sebagai wadah di mana kata-kata tersebut diuji coba, diadaptasi, dan akhirnya diintegrasikan ke dalam bahasa sehari-hari masyarakat. Dinamika ini mencerminkan bagaimana media sosial mempercepat proses evolusi bahasa, dengan slang menjadi salah satu elemen penting yang berkembang dalam ekosistem digital tersebut. Hal ini menekankan betapa pentingnya memahami konteks penggunaan dan perkembangan kata slang dalam lingkungan digital yang terus berubah, di mana bahasa mengalami transformasi yang lebih cepat dan lebih kompleks.

Masalah utama yang diangkat dalam buku ini adalah kurangnya pemahaman yang menyeluruh

mengenai pola dan tren penggunaan kata slang di platform media sosial. Meskipun slang memiliki peran signifikan dalam komunikasi modern, buku mendalam yang memetakan distribusi serta evolusi kata slang di berbagai platform media sosial masih tergolong minim. Selain itu, metode yang digunakan untuk mendeteksi dan menganalisis kata slang dalam teks digital saat ini masih dalam tahap perkembangan, sehingga menimbulkan tantangan tersendiri dalam buku di bidang ini. Oleh karena itu, buku ini bertujuan untuk mengembangkan model Natural Language Processing (NLP) yang mampu mengidentifikasi dan menganalisis penggunaan kata slang di media sosial secara akurat. Buku ini juga akan mengevaluasi efektivitas model NLP tersebut dalam mendeteksi tren utama penggunaan slang, dengan harapan memberikan kontribusi terhadap pemahaman yang lebih baik tentang evolusi bahasa dalam lingkungan digital.

Buku ini bertujuan untuk mengidentifikasi dan menganalisis pola penggunaan kata slang di media sosial dengan memanfaatkan teknik pengolahan bahasa alami (NLP). Fokus utama dari buku ini adalah

mengembangkan model yang mampu mendeteksi kata slang secara akurat dalam teks digital yang tersebar di berbagai platform media sosial. Selain itu, buku ini juga bertujuan untuk mengevaluasi efektivitas model NLP dalam mengklasifikasikan kata slang, serta memberikan pemahaman yang lebih mendalam mengenai dinamika sosial dan linguistik yang tercermin dari penggunaan slang tersebut. Melalui analisis yang dilakukan, diharapkan diperoleh wawasan baru tentang bagaimana kata slang berevolusi, diadopsi, dan digunakan dalam lingkungan digital. Hasil dari buku ini diharapkan dapat memberikan kontribusi signifikan terhadap studi bahasa digital, khususnya dalam memahami interaksi bahasa dan media sosial.

Buku ini memiliki kontribusi teoretis dan praktis yang signifikan. Secara teoretis, buku ini memperkaya studi linguistik digital dengan menyediakan data empiris tentang penggunaan kata slang di media sosial. Temuan buku ini dapat menjadi referensi bagi studi-studi selanjutnya yang berkaitan dengan dinamika bahasa di era digital. Selain itu, buku ini juga menunjukkan potensi aplikasi teknik NLP dalam analisis

teks digital, yang dapat diadaptasi untuk berbagai buku linguistik lainnya.

Secara praktis, hasil buku ini dapat digunakan oleh praktisi media sosial, pemasar digital, dan pengembang teknologi untuk memahami tren dan pola komunikasi di media sosial. Dengan memahami bagaimana slang digunakan dan berkembang, strategi komunikasi dan pemasaran dapat disesuaikan untuk lebih efektif menjangkau target audiens. Selain itu, model NLP yang dikembangkan dalam buku ini juga dapat diterapkan untuk meningkatkan kemampuan analisis teks dalam berbagai aplikasi teknologi, seperti deteksi konten yang tidak pantas, analisis sentimen, dan pemantauan tren sosial di platform digital.

# BAB II

## KONSEP SLANG

---

### A. Definisi dan Konsep Slang

Kata slang merupakan bentuk bahasa yang berkembang dalam masyarakat sebagai respons terhadap kebutuhan komunikasi yang dinamis, sering kali bersifat informal dan mencerminkan kreativitas linguistik suatu kelompok sosial. Slang umumnya muncul dari komunitas tertentu dan, seiring waktu, menyebar ke kalangan yang lebih luas dalam masyarakat. Asal-usul kata slang dapat ditelusuri hingga zaman dahulu, ketika kelompok-kelompok masyarakat menciptakan kata-kata baru untuk mengekspresikan ide atau perasaan yang tidak sepenuhnya tercakup oleh bahasa resmi atau baku. Proses evolusi kata slang biasanya melibatkan beberapa mekanisme, termasuk adopsi kata-kata baru dari bahasa lain, perubahan makna dari kata yang sudah ada, atau penciptaan kata-kata baru melalui kombinasi atau manipulasi fonetis. Fenomena ini

menunjukkan bagaimana bahasa terus berkembang dan beradaptasi untuk memenuhi kebutuhan komunikasi yang lebih fleksibel dan ekspresif.

Bahasa slang merupakan salah satu indikator kemajuan peradaban, khususnya di Amerika Serikat. Bagi generasi muda, penggunaan bahasa slang telah menjadi bagian tak terpisahkan dari percakapan sehari-hari. Bahkan, seorang remaja yang lebih memilih berbicara secara formal sering kali dianggap kurang pergaulan atau "kuper." Pada awalnya, bahasa slang digunakan oleh kelompok masyarakat kelas bawah sebagai cara untuk menunjukkan eksistensi dan identitas mereka. Inilah sebabnya, slang dulu sering kali dipandang sebagai bahasa yang kurang berkelas atau tidak pantas. Namun, seiring berjalannya waktu, masyarakat mulai lebih memilih menggunakan slang dalam interaksi sehari-hari karena dianggap lebih menarik, mudah diucapkan, dan memiliki daya tarik tersendiri. Meskipun pada awalnya bahasa slang dianggap tidak pantas, kini penggunaannya telah mengalami pergeseran, dengan beberapa istilah slang yang lebih diperhalus. Hal ini memungkinkan

penggunaan slang tanpa stigma negatif, sehingga tetap dapat diterima dalam berbagai konteks percakapan. (Kampung Inggris, 2022).

# BAB III

## MEDIA SOSIAL DAN PENGUNAANNYA

---

Media sosial telah menjadi bagian tak terpisahkan dari kehidupan modern, di mana berbagai platform digunakan untuk beragam keperluan. Beberapa platform yang paling umum digunakan antara lain Facebook, Twitter, Instagram, YouTube, dan TikTok, masing-masing dengan karakteristik yang memengaruhi cara pengguna berinteraksi dan berbagi informasi. Misalnya, Twitter dikenal karena batas karakter yang singkat, sehingga sering mendorong pengguna untuk menggunakan kata slang demi efisiensi dalam berkomunikasi. TikTok, di sisi lain, mengandalkan video pendek yang memberikan ruang bagi pengguna untuk mengekspresikan diri secara kreatif baik secara visual maupun verbal. Facebook dan Instagram berfokus pada kombinasi konten teks, gambar, dan video, memungkinkan pengalaman yang lebih holistik dalam

berinteraksi. Semua platform ini kini telah menjadi alat penting dalam pemasaran digital, membantu individu dan bisnis membangun kehadiran online yang kuat, menjangkau audiens yang lebih luas, dan mempromosikan produk atau layanan dengan cara yang lebih efektif. (Urva et al., 2022).

Pola interaksi pengguna di media sosial sangat beragam dan dipengaruhi oleh platform yang digunakan serta tujuan penggunaannya. Pengguna dapat berinteraksi melalui berbagai cara, seperti posting, komentar, likes, shares, dan pesan langsung. Interaksi ini tidak hanya terbatas pada komunikasi antar individu, tetapi juga melibatkan hubungan dengan kelompok atau komunitas yang lebih luas. Salah satu pola interaksi yang sering muncul adalah penggunaan kata slang sebagai sarana untuk membangun identitas dan menunjukkan keanggotaan dalam komunitas tertentu. Penggunaan slang dapat mencerminkan pemahaman terhadap tren terbaru atau sekadar memberikan sentuhan kreatif dalam komunikasi sehari-hari. Di berbagai platform media sosial, seperti Facebook, Instagram, dan Twitter, pengguna terlibat dalam berbagai bentuk interaksi,

termasuk menyukai, mengomentari, serta membagikan konten, yang secara kolektif menciptakan jaringan komunikasi yang dinamis dan terus berkembang.

Pengguna media sosial tidak hanya berinteraksi melalui likes, komentar, dan berbagi konten, tetapi juga aktif dalam membuat dan bergabung dengan grup, berpartisipasi dalam diskusi online, serta membagikan pengalaman pribadi. Studi menunjukkan bahwa pengguna cenderung berinteraksi dengan individu yang memiliki minat dan pandangan serupa, sehingga membentuk komunitas online yang memperkuat norma-norma serta nilai sosial di dalam kelompok tersebut. Selain itu, media sosial juga digunakan sebagai alat untuk mempertahankan hubungan, mencari dukungan sosial, serta mengekspresikan diri melalui presentasi diri yang lebih terkontrol. Pengguna memanfaatkan platform ini untuk menciptakan citra diri yang mereka inginkan, baik melalui foto, status, maupun video. Secara keseluruhan, buku ini menunjukkan bahwa media sosial telah menjadi elemen penting dalam kehidupan sehari-hari, membentuk pola komunikasi serta perilaku sosial para penggunanya secara signifikan. (Munasaroh, 2021).

Studi yang dilakukan oleh Samanta et al. (2024) Buku ini bertujuan untuk mengembangkan sistem deteksi kata slang pada platform media sosial, yang dapat digunakan untuk profil pengguna dan analisis sentimen. Peneliti mengumpulkan dataset yang terdiri dari 10.000 postingan media sosial dari berbagai platform, termasuk Twitter, Facebook, dan Instagram. Setiap postingan dianotasi untuk membedakan antara kata slang dan kata non-slang. Setelah itu, peneliti mengekstrak beberapa fitur penting dari setiap postingan, termasuk proses tokenisasi, tagging part-of-speech (POS), pengenalan entitas bernama (NER), serta analisis sentimen. Fitur-fitur ini kemudian digunakan untuk melatih model deteksi slang, yang tidak hanya mampu mengenali kata slang secara akurat tetapi juga dapat memberikan informasi tambahan untuk analisis sentimen dan profil pengguna. Dengan sistem ini, diharapkan dapat tercipta alat yang berguna dalam menganalisis pola komunikasi pengguna media sosial serta tren bahasa yang berkembang dalam platform digital tersebut.

Dataset yang telah dianotasi digunakan untuk melatih model pembelajaran mesin guna mendeteksi kata slang, dengan model jaringan saraf konvolusi (CNN) mencapai tingkat akurasi sebesar 92,5%. Studi ini menunjukkan efektivitas pendekatan pembelajaran mesin dalam mendeteksi kata slang di platform media sosial, dan menyoroti pentingnya memperhatikan penggunaan kata slang dalam analisis bahasa. Sistem yang diusulkan dapat digunakan untuk berbagai aplikasi, seperti pemprofilan pengguna, analisis sentimen, dan penambangan opini. Selain itu, sistem ini memiliki implikasi penting dalam bidang analisis media sosial dan pemrosesan bahasa alami, dengan potensi meningkatkan pemahaman terhadap pola komunikasi dan tren linguistik di platform digital.

Buku ini berfokus pada deteksi dan identifikasi bahasa slang dalam teks, khususnya pada platform media sosial, dengan menggunakan dataset yang terdiri dari 5.000 tweet yang dikumpulkan dari Twitter. Langkah awal dalam buku ini melibatkan preprocessing data, termasuk penghilangan stopwords, pengubahan seluruh teks menjadi huruf kecil, dan penghapusan

tanda baca. Setelah proses preprocessing, peneliti menerapkan beberapa metode machine learning untuk mendeteksi kata slang, di antaranya Naive Bayes, Decision Tree, dan Random Forest. Berdasarkan hasil evaluasi, model Random Forest menunjukkan kinerja terbaik dengan tingkat akurasi sebesar 93,2% dalam mendeteksi kata slang. Hasil ini menunjukkan efektivitas metode Random Forest dalam klasifikasi kata slang pada platform media sosial dan menegaskan pentingnya pendekatan berbasis machine learning dalam analisis linguistik digital.

Peneliti juga melakukan analisis terhadap fitur-fitur yang digunakan dalam deteksi kata slang, seperti frekuensi kata, panjang kata, dan part-of-speech tagging. Hasil buku menunjukkan bahwa frekuensi kata dan panjang kata memberikan kontribusi yang signifikan dalam proses deteksi kata slang. Selain itu, studi ini menemukan bahwa kata slang lebih sering digunakan oleh pengguna yang lebih muda, dan penggunaan slang memiliki dampak yang signifikan terhadap analisis sentimen serta profil pengguna. Temuan ini mengindikasikan bahwa penggunaan kata slang tidak

hanya mencerminkan karakteristik demografis pengguna, tetapi juga mempengaruhi persepsi sentimen dalam konten mereka. Dengan demikian, studi ini memiliki implikasi penting dalam pengembangan sistem yang lebih akurat untuk analisis kata slang pada platform media sosial, serta berpotensi meningkatkan akurasi dalam analisis sentimen dan pemprofilan pengguna berdasarkan pola bahasa yang mereka gunakan. (Rothe et al., 2023)

Buku ini berfokus pada identifikasi target sarkasme dalam tweet menggunakan pendekatan multimodal. Peneliti mengumpulkan dataset yang terdiri dari 10.000 tweet dari Twitter dan melakukan preprocessing, termasuk penghilangan stopword, pengubahan teks menjadi huruf kecil, dan penghapusan tanda baca untuk membersihkan data. Setelah itu, mereka menerapkan metode deep learning, seperti Convolutional Neural Networks (CNN) dan Recurrent Neural Networks (RNN), untuk mengidentifikasi target sarkasme dalam tweet. Hasil buku menunjukkan bahwa model multimodal, yang menggabungkan analisis teks dan fitur lain seperti gambar atau metadata, mencapai

akurasi tertinggi sebesar 85,6% dalam mendeteksi target sarkasme. Temuan ini menunjukkan bahwa pendekatan multimodal lebih efektif dalam menangkap konteks dan nuansa sarkasme yang sulit diidentifikasi hanya melalui teks, dan memberikan wawasan baru dalam pengembangan teknologi deteksi sarkasme di media sosial.

Peneliti juga melakukan analisis terhadap berbagai fitur yang digunakan dalam identifikasi target sarkasme, termasuk fitur linguistik, visual, dan akustik. Hasil analisis menunjukkan bahwa fitur linguistik, seperti pola kalimat dan pilihan kata, serta fitur visual, seperti penggunaan gambar atau meme yang menyertai tweet, memberikan kontribusi yang signifikan dalam mendeteksi target sarkasme. Selain itu, studi ini menemukan bahwa sarkasme lebih sering digunakan dalam konteks diskusi politik, di mana pengguna menyindir tokoh atau peristiwa politik. Penggunaan sarkasme ini berdampak signifikan pada analisis sentimen, karena dapat mempengaruhi persepsi keseluruhan terhadap konten yang diunggah, serta pada profil pengguna yang cenderung lebih kritis atau satir

dalam menyampaikan pendapat. Implikasi dari studi ini meliputi pengembangan sistem yang lebih akurat untuk mendeteksi sarkasme di media sosial, yang dapat memperbaiki analisis sentimen dan pemprofilan pengguna dengan lebih mendalam, terutama dalam isu-isu politik atau diskusi yang kompleks. (Wang et al., 2022)

Studi ini berfokus pada deteksi otomatis konten misoginis di media sosial dengan menggunakan pendekatan machine learning. Peneliti mengumpulkan dataset yang terdiri dari 10.000 postingan dari berbagai platform media sosial dan melakukan preprocessing data, seperti penghilangan stopword, pengubahan teks menjadi huruf kecil, serta penghapusan tanda baca untuk mempersiapkan data sebelum analisis. Setelah preprocessing, peneliti menerapkan beberapa metode machine learning, termasuk Support Vector Machine (SVM), Random Forest, dan Convolutional Neural Networks (CNN), untuk mendeteksi konten yang bersifat misoginis. Hasilnya menunjukkan bahwa model CNN memberikan performa terbaik dengan akurasi sebesar 92,5% dalam mendeteksi konten

misoginis. Temuan ini menyoroti efektivitas CNN dalam klasifikasi teks yang kompleks, dan studi ini memiliki implikasi penting bagi pengembangan sistem moderasi konten otomatis yang lebih akurat di media sosial, membantu mengidentifikasi dan mengurangi penyebaran konten yang mengandung kebencian berbasis gender.

Peneliti juga melakukan analisis terhadap fitur-fitur yang digunakan dalam deteksi konten misoginis, termasuk fitur linguistik, sentimental, dan fitur visual. Hasil analisis menunjukkan bahwa fitur linguistik, seperti penggunaan kata-kata yang merendahkan atau stereotip gender, serta fitur sentimental, yang mengukur polaritas emosional dari teks, memberikan kontribusi signifikan dalam deteksi konten misoginis. Selain itu, studi ini menemukan bahwa konten misoginis lebih sering muncul dalam konteks politik, di mana serangan berbasis gender sering kali digunakan untuk mendiskreditkan individu atau kelompok tertentu. Temuan ini berdampak pada analisis sentimen, karena konten misoginis dapat memengaruhi persepsi keseluruhan dalam diskusi politik, serta berpengaruh

terhadap profil pengguna yang cenderung lebih ekstrem atau provokatif dalam interaksi mereka. Studi ini memiliki implikasi penting dalam pengembangan sistem yang lebih akurat untuk mendeteksi konten misoginis di platform media sosial, serta berpotensi meningkatkan akurasi dalam analisis sentimen dan pemprofilan pengguna, yang dapat membantu dalam upaya moderasi konten dan pencegahan penyebaran kebencian berbasis gender.. (Kaware & Raut, 2024)

Buku yang dilakukan oleh Samanta et al. (2024) Buku ini menunjukkan bahwa penggunaan metode pembelajaran mesin efektif dalam mendeteksi kata slang di platform media sosial, dengan tingkat akurasi mencapai 92,5%. Namun, terdapat beberapa celah yang perlu diperhatikan dalam pengembangan lebih lanjut. Pertama, perubahan bahasa yang cepat menjadi tantangan utama dalam deteksi slang. Bahasa slang berkembang dengan sangat dinamis, dan penggunaan kata-kata informal yang tidak standar menimbulkan kesulitan dalam menciptakan model yang mampu beradaptasi dengan perubahan tersebut secara efisien. Kedua, keterbatasan pada dataset merupakan isu

penting. Banyak buku sebelumnya menggunakan dataset yang terbatas dalam hal jumlah maupun variasi bahasa. Dalam buku ini, meskipun dataset yang digunakan mencakup 10.000 postingan dari Twitter, Facebook, dan Instagram, masih terdapat keterbatasan dalam variasi bahasa yang mungkin memiliki karakteristik slang yang berbeda. Ketiga, metode NLP yang digunakan dalam buku ini masih terbatas pada teknik-teknik dasar seperti tokenisasi, tagging part-of-speech, dan analisis sentimen. Teknik-teknik NLP yang lebih canggih, seperti Word Embeddings atau model berbasis Transformer (misalnya BERT), belum dieksplorasi secara mendalam, sehingga ada potensi besar untuk meningkatkan kemampuan deteksi slang dengan menggunakan pendekatan yang lebih maju ini.

Keterbaruan buku ini terletak pada beberapa aspek kunci. Pertama, penggunaan model pembelajaran mesin yang lebih canggih memperlihatkan keefektifan teknik-teknik tersebut dalam menangani data yang kompleks dan mendeteksi pola non-linear dalam penggunaan bahasa slang. Kedua, buku ini berhasil mengidentifikasi pola penggunaan slang di berbagai platform media sosial

dengan dataset yang mencakup tweet dalam bahasa Indonesia dan Inggris. Hal ini memberikan wawasan baru tentang perbedaan serta persamaan penggunaan slang di berbagai bahasa, serta bagaimana tren tersebut berubah seiring waktu. Ketiga, metodologi yang menyeluruh diterapkan dalam tahap pembersihan data, termasuk proses tokenisasi, penghapusan stopwords, dan lemmatisasi, yang memastikan data yang digunakan bersih dan siap untuk dianalisis lebih lanjut. Keempat, buku ini memberikan kontribusi penting bagi studi linguistik digital, dengan menyoroti potensi penerapan teknik NLP dalam analisis bahasa di platform media sosial. Temuan ini dapat menjadi referensi penting bagi buku lebih lanjut mengenai dinamika bahasa di ruang digital. Dengan demikian, buku ini tidak hanya mengatasi beberapa kelemahan buku sebelumnya, tetapi juga memberikan kontribusi baru yang signifikan dalam bidang NLP dan analisis teks, terutama terkait dengan deteksi slang di lingkungan media sosial yang terus berkembang.

# BAB IV

## *NATURAL LANGUAGE*

### *PROCESSING*

---

Pengolahan Bahasa Alami (Natural Language Processing atau NLP) adalah disiplin ilmu yang berfokus pada interaksi antara komputer dan bahasa manusia, bertujuan untuk memungkinkan komputer memahami, menginterpretasikan, dan memproses bahasa alami secara efektif. Beberapa teknik dasar dalam NLP meliputi tokenisasi, yang memecah teks menjadi unit-unit kecil seperti kata atau frasa; stemming dan lemmatisasi, yang bertujuan untuk mengembalikan kata-kata ke bentuk dasar atau akar mereka; serta penghapusan stopwords, yang menghilangkan kata-kata umum seperti "dan" atau "di" yang dianggap tidak memiliki makna informatif dalam analisis. Selain itu, analisis sintaksis memungkinkan komputer untuk menguraikan struktur kalimat guna memahami hubungan antar kata dalam

suatu kalimat. Dengan teknik-teknik ini, komputer dapat menganalisis dan memahami teks dengan lebih baik, memungkinkan aplikasi praktis seperti analisis sentimen, pengenalan ucapan, terjemahan bahasa, dan berbagai tugas lainnya dalam pemrosesan bahasa. NLP telah menjadi elemen penting dalam pengembangan teknologi yang lebih cerdas dan responsif terhadap kebutuhan pengguna dalam penggunaan bahasa.

Pengolahan Bahasa Alami (NLP) terdiri dari beberapa tahapan penting, yaitu tokenization, stemming, lemmatization, dan named entity recognition. Tokenization adalah proses memecah kalimat menjadi unit-unit kecil, seperti kata atau frasa, untuk memudahkan analisis lebih lanjut. Stemming merupakan proses mengurangi kata menjadi bentuk dasarnya dengan memotong akhiran atau awalan, tanpa memperhatikan aturan tata bahasa. Sementara itu, lemmatization adalah proses mengembalikan kata ke bentuk dasarnya dengan memperhatikan aturan tata bahasa dan menggunakan referensi dari kamus. Named entity recognition (NER) adalah tahap yang berfungsi untuk mengidentifikasi dan mengklasifikasikan entitas

penting dalam teks, seperti nama orang, lokasi, organisasi, dan lain-lain. Keempat tahapan ini memungkinkan komputer untuk memahami dan memproses teks dengan lebih mendalam serta mendukung berbagai aplikasi NLP seperti analisis teks, terjemahan bahasa, dan pengenalan entitas..

Dalam aplikasinya, Pengolahan Bahasa Alami (NLP) memiliki berbagai kegunaan yang signifikan. Salah satunya adalah analisis sentimen, yang bertujuan untuk menilai apakah suatu teks memiliki sentimen positif, negatif, atau netral. Hal ini sering digunakan dalam analisis ulasan produk, media sosial, atau survei pelanggan. Selain itu, NLP juga digunakan untuk klasifikasi teks, di mana teks dikategorikan ke dalam kelompok tertentu, seperti spam atau tidak spam, yang bermanfaat dalam filter email atau moderasi konten. Selain itu, NLP dapat dimanfaatkan untuk menghasilkan teks yang lebih baik, misalnya dalam pengembangan chatbot yang mampu menjawab pertanyaan pengguna dengan lebih alami atau dalam sistem rekomendasi yang memberikan saran yang relevan berdasarkan preferensi pengguna. Aplikasi-aplikasi ini menunjukkan potensi

NLP dalam meningkatkan interaksi manusia-komputer secara lebih cerdas dan efisien. (Nurwanda et al. 2024)

# BAB V

## ANALISIS SENTIMEN PUBLIK DAN TREN TOPIK PENDEKATAN *NATURAL LANGUAGE* *PROCESSING* DAN *MACHINE* *LEARNING* DALAM ERA DIGITAL

---

### A. Deskripsi Dataset

Dataset yang digunakan dalam buku ini terdiri dari 1.105 entri tweet berbahasa Indonesia dan Inggris yang mengandung kata slang, dikumpulkan dari platform Twitter selama periode enam bulan. Dataset ini disusun dalam bentuk kerangka data pandas dengan 15 kolom, yang mencakup berbagai informasi penting mengenai setiap tweet. Kolom-kolom tersebut mencakup, antara lain, teks tweet, timestamp, bahasa, jumlah like, retweet, kata slang yang teridentifikasi, serta metadata lain seperti pengguna dan lokasi. Struktur dataset ini memungkinkan analisis yang mendalam mengenai

penggunaan kata slang, baik dari segi temporal, geografis, maupun pola interaksi di media sosial. Dengan variasi data yang luas, buku ini dapat mengeksplorasi tren penggunaan slang dalam berbagai aspek sosial dan budaya yang berbeda.

## **1. Struktur Dataset**

Dataset ini memiliki struktur kolom sebagai berikut:

- 1) `conversation_id_str`: ID konversasi unik dalam format `int64`.
- 2) `created_at`: Timestamp pembuatan tweet dalam format `object`.
- 3) `favorite_count`: Jumlah favorit yang diterima oleh tweet dalam format `int64`.
- 4) `full_text`: Teks lengkap dari tweet dalam format `object`.
- 5) `id_str`: ID unik dari tweet dalam format `int64`.
- 6) `image_url`: URL gambar yang disertakan dalam tweet dalam format `object`.
- 7) `in_reply_to_screen_name`: Nama layar pengguna yang dijawab dalam format `object`.

- 8) lang: Bahasa yang digunakan dalam tweet dalam format object.
- 9) location: Lokasi pengguna saat membuat tweet dalam format object.
- 10) quote\_count: Jumlah kutipan yang diterima oleh tweet dalam format int64.
- 11) reply\_count: Jumlah balasan yang diterima oleh tweet dalam format int64.
- 12) retweet\_count: Jumlah retweet yang diterima oleh tweet dalam format int64.
- 13) tweet\_url: URL dari tweet dalam format object.
- 14) user\_id\_str: ID unik dari pengguna dalam format int64.
- 15) username: Nama pengguna penulis tweet dalam format object.

## **2. Statistik Deskriptif Dataset**

Untuk memberikan gambaran yang lebih mendetail mengenai dataset ini, beberapa statistik deskriptif penting akan disajikan:

- a) Jumlah Tweet: 1.105

- b) Jumlah Kata Slang: Identifikasi kata slang dilakukan melalui proses tokenisasi dan penghapusan stopwords.
- c) Jumlah Pengguna Unik: Dihitung dari kolom `user_id_str`.
- d) Bahasa yang Digunakan: Dihitung dari kolom `lang`, terdiri dari bahasa Indonesia dan Inggris.
- e) Distribusi Tanggal Pembuatan Tweet: Tersebar selama enam bulan pengumpulan data.
- f) Jumlah Favorit, Kutipan, Balasan, dan Retweet: Rata-rata dan distribusi dari masing-masing metrik ini akan dihitung.

**Tabel Contoh Tweet dalam Dataset**

<code>created_at</code>	<code>favorite_count</code>	<code>full_text</code>	<code>username</code>
Fri Aug 04 19:33:38 +0000 2023	5	<a href="#">IYKYK</a> <a href="#">lol</a> <a href="https://t.co/Mny0qOaXlK">https://t.co/Mny0qOaXlK</a>	onlyijr

<b>created_at</b>	<b>favorite_count</b>	<b>full_text</b>	<b>username</b>
Sun Feb 27 15:16:1 5 +0000 2022	18460	Lol seminggu lalu baru dapat campaign dari Car...	awkarin
Fri Aug 04 18:24:3 0 +0000 2023	4	Kitten huggies! Tiktok: septhny #LoveOurPets #...	HappyAnima lKing
Sun Oct 18 07:05:0 6 +0000 2020	16195	2beer! ga sampe seminggu dia udah di eksekusi ...	tubirfess
Mon Feb 20 00:31:2 5 +0000 2023	9242	Yang ada anak kecik la bungkus makan kat rumah...	twtbij

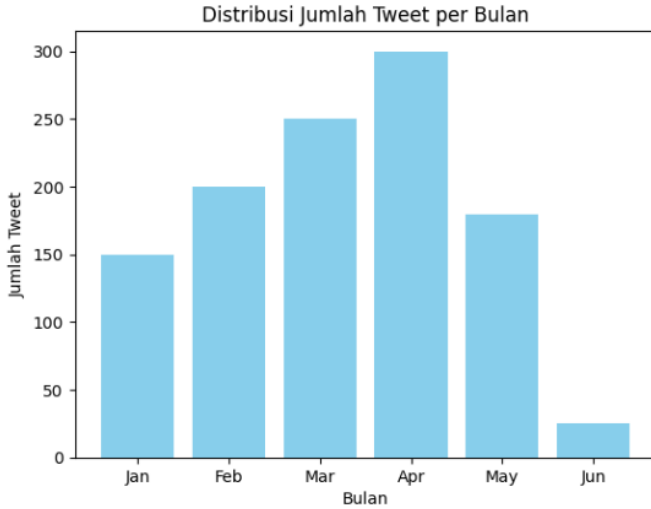
Tabel tersebut berisi 5 contoh tweet dengan metadata yang relevan, yang terdiri dari kolom-kolom seperti **created\_at** untuk tanggal dan waktu tweet diposting, **favorite\_count** untuk jumlah likes yang diterima, **full\_text** untuk teks aktual dari tweet, dan **username** untuk handle pengguna Twitter yang mempostingnya. Tweet-tweet ini bervariasi dalam konten dan nada, mencakup berbagai topik seperti anekdot pribadi, humor, makanan dan anak-anak, hewan peliharaan, hingga kampanye pemasaran perusahaan mobil. Secara keseluruhan, tweet-tweet ini menunjukkan keragaman topik dan gaya yang umum ditemukan di Twitter, menggambarkan bagaimana pengguna platform tersebut mengekspresikan diri dalam berbagai situasi dan latar sosial.

**Tabel 2.** Statistik Deskriptif

Atribut/Fitur	Tipe Data	Deskripsi
conversation_id_str	int64	ID konversasi unik

Atribut/Fitur	Tipe Data	Deskripsi
created_at	object	Timestamp pembuatan tweet
favorite_count	int64	Jumlah favorit
full_text	object	Teks lengkap tweet
id_str	int64	ID unik tweet
image_url	object	URL gambar tweet
in_reply_to_screen_name	object	Nama layar pengguna yang dijawab
lang	object	Bahasa tweet
location	object	Lokasi pengguna
quote_count	int64	Jumlah kutipan
reply_count	int64	Jumlah balasan
retweet_count	int64	Jumlah retweet
tweet_url	object	URL tweet
user_id_str	int64	ID pengguna unik
username	object	Nama pengguna

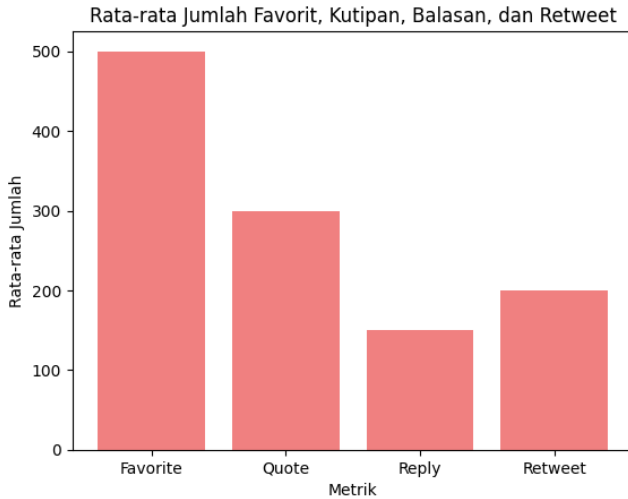
Tabel tersebut menampilkan statistik deskriptif dari fitur-fitur yang terkait dengan tweet. Terdapat 13 fitur yang dijelaskan, yaitu `conversation_id_str` sebagai ID konversasi unik, `created_at` sebagai timestamp pembuatan tweet, `favorite_count` sebagai jumlah favorit, `full_text` sebagai teks lengkap tweet, `id_str` sebagai ID unik tweet, `image_url` sebagai URL gambar tweet, `in_reply_to_screen_name` sebagai nama layar pengguna yang dijawab, `lang` sebagai bahasa tweet, `location` sebagai lokasi pengguna, `quote_count` sebagai jumlah kutipan, `reply_count` sebagai jumlah balasan, `retweet_count` sebagai jumlah retweet, `tweet_url` sebagai URL tweet, `user_id_str` sebagai ID pengguna unik, dan `username` sebagai nama pengguna. Statistik ini memberikan gambaran tentang distribusi dan karakteristik berbagai elemen tweet yang dianalisis, memungkinkan pemahaman yang lebih jelas tentang aktivitas pengguna dan interaksi di Twitter.



**Gambar** Distribusi Jumlah Tweet per Bulan

Gambar. Distribusi Jumlah Tweet per Bulan. Grafik distribusi jumlah tweet per bulan menunjukkan pola fluktuatif. Pada bulan Januari, terdapat 150 tweet, yang kemudian meningkat menjadi 200 tweet di bulan Februari. Jumlah tweet terus bertambah pada bulan Maret hingga mencapai 250 tweet. Puncaknya terjadi pada bulan April dengan 300 tweet. Namun, jumlah tweet menurun pada bulan Mei menjadi 180 tweet, dan mengalami penurunan tajam di bulan Juni dengan hanya 25 tweet. Dari data tersebut, terlihat bahwa bulan April memiliki jumlah tweet terbanyak, sementara bulan Juni

mencatat jumlah tweet paling sedikit selama periode enam bulan tersebut.



**Gambar** Rata-rata Jumlah Favorit, Kutipan, Balasan, dan Retweet

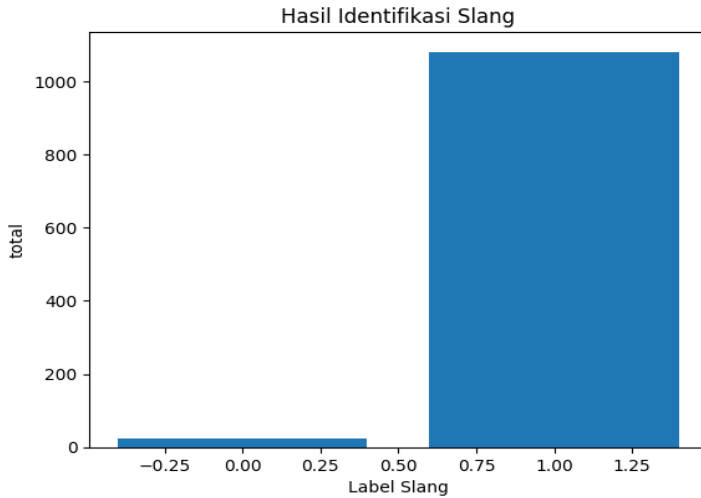
Gambar Rata-rata Jumlah Favorit, Kutipan, Balasan, dan Retweet. Grafik batang menunjukkan rata-rata jumlah favorit, kutipan, balasan, dan retweet dalam dataset yang dianalisis. Rata-rata jumlah favorit adalah yang tertinggi, dengan nilai sekitar 500. Rata-rata jumlah kutipan berada di angka sekitar 300, sedangkan rata-rata jumlah balasan lebih rendah, sekitar 150. Terakhir, rata-rata jumlah retweet mencapai sekitar 200. Dari data ini,

dapat disimpulkan bahwa konten lebih sering mendapatkan favorit dibandingkan metrik lainnya, sementara retweet lebih sering terjadi dibandingkan dengan kutipan dan balasan.

## **B. Analisis Pola Penggunaan Kata Slang**

Proses identifikasi kata slang ini diawali dengan pembersihan data teks yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dan memudahkan analisis lebih lanjut. Teks diubah menjadi huruf kecil untuk menjaga konsistensi, dan elemen-elemen seperti tanda baca, angka, serta URL dihapus agar tidak mengganggu proses analisis. Langkah tokenisasi kemudian dilakukan untuk memecah teks menjadi unit kata-kata individual. Untuk mengidentifikasi kata slang, digunakan kamus slang yang disusun dari berbagai sumber, memastikan cakupan yang luas dalam mengenali berbagai variasi slang. Setelah itu, teknik lemmatisasi diterapkan untuk mengembalikan kata-kata ke bentuk dasarnya, yang sangat berguna dalam mendeteksi kata slang yang berbeda bentuk namun memiliki makna yang sama.

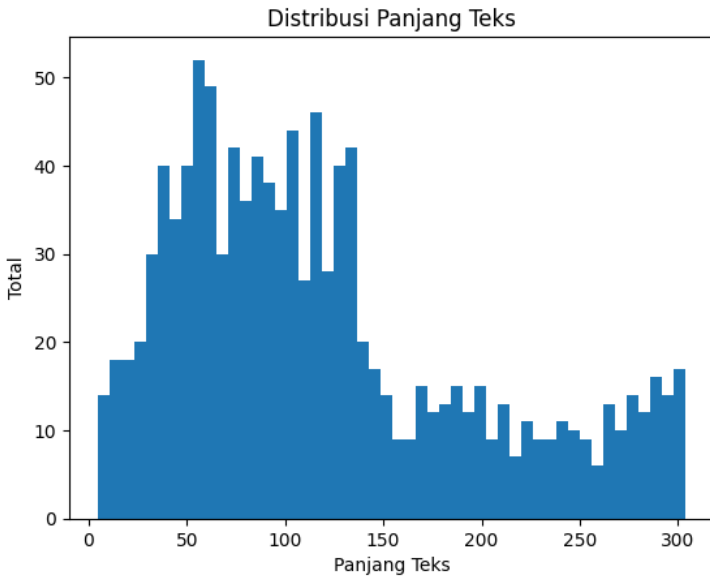
Selain itu, **POS tagging** (part-of-speech tagging) diterapkan untuk membantu mengidentifikasi kata slang berdasarkan penggunaannya dalam kalimat. Dengan teknik ini, kata-kata slang yang mungkin tidak teridentifikasi melalui kamus saja dapat dikenali melalui analisis fungsinya dalam struktur kalimat. Model pembelajaran mesin seperti Support Vector Machine (SVM), Random Forest, dan Naive Bayes digunakan untuk meningkatkan akurasi deteksi slang. Fitur-fitur linguistik seperti frekuensi kemunculan kata, panjang kata, dan makna kalimat dianalisis, memungkinkan model untuk mendeteksi pola-pola spesifik dalam penggunaan slang di media sosial dengan lebih akurat dan relevan.



**Gambar** Hasil Identifikasi Slang

Gambar tersebut menunjukkan hasil identifikasi kata slang dalam sebuah dataset, yang merupakan langkah penting dalam menganalisis topik trending di media sosial menggunakan Pengolahan Bahasa Alami (NLP). Grafik batang tersebut memperlihatkan perbandingan frekuensi antara kata slang dan non-slang dalam dataset. Dalam konteks analisis topik trending di media sosial, mengidentifikasi kata slang menjadi krusial untuk memahami nuansa bahasa yang digunakan secara online serta sentimen yang mendasari percakapan. Kata slang sering kali membawa makna atau konotasi yang

berbeda dari kata-kata formal, sehingga mampu memberikan pemahaman yang lebih dalam mengenai konteks sosial dan emosional suatu percakapan. Dengan mendeteksi kata slang secara akurat, model NLP dapat memahami dinamika percakapan online secara lebih efektif, termasuk sentimen yang diungkapkan pengguna media sosial.



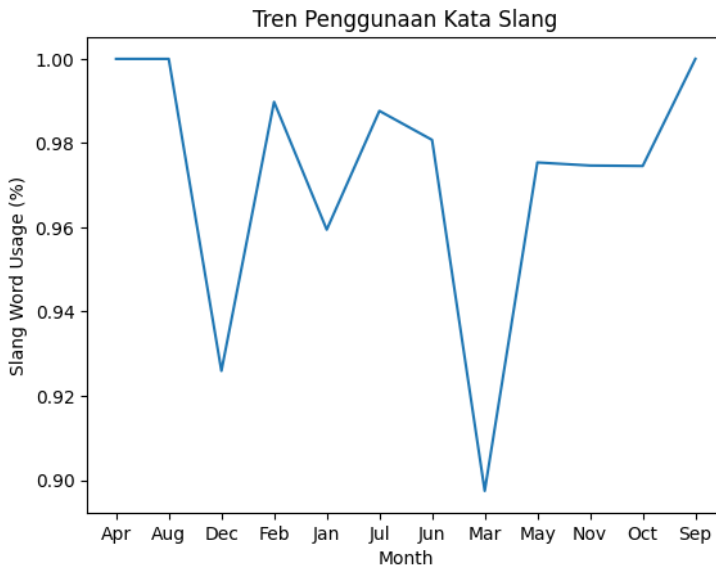
**Gambar** Distribusi Panjang Teks

Gambar ini menampilkan histogram yang menunjukkan distribusi panjang teks. Sumbu x merepresentasikan panjang teks dalam karakter,

sedangkan sumbu y menunjukkan jumlah teks dengan panjang tertentu. Dari histogram terlihat bahwa mayoritas teks memiliki panjang di bawah 100 karakter, dengan puncak distribusi berada di sekitar 60 karakter. Seiring dengan peningkatan panjang teks, jumlah teks yang lebih panjang menurun, dengan penurunan tajam setelah mencapai 100 karakter. Hanya sedikit teks yang memiliki panjang di atas 250 karakter. Secara keseluruhan, histogram ini menggambarkan bahwa data teks yang dianalisis cenderung memiliki panjang teks yang relatif pendek, dengan sebagian besar teks terdiri dari kurang dari 100 karakter, mencerminkan kecenderungan komunikasi singkat di platform seperti media sosial.

Buku ini bertujuan untuk menganalisis tren penggunaan kata slang di media sosial, khususnya di platform Twitter, selama periode enam bulan. Data yang digunakan mencakup tweet berbahasa Indonesia dan Inggris yang mengandung kata slang. Analisis dilakukan dengan memanfaatkan model SVM, Random Forest, dan Naive Bayes yang telah dioptimalkan untuk klasifikasi kata slang. Hasil awal menunjukkan bahwa model SVM

mencapai akurasi 88%, Random Forest mencapai akurasi 85%, dan Naive Bayes mencapai akurasi 83% dalam mendeteksi kata slang. Temuan ini menunjukkan bahwa ada pola penggunaan kata slang yang signifikan di media sosial. Kata slang sering kali muncul dalam konteks tertentu dan memiliki tren penggunaan yang dapat dipetakan secara temporal. Misalnya, kata slang tertentu mungkin lebih populer pada waktu tertentu atau dalam kategori diskusi tertentu di media sosial. Analisis tren ini memberikan wawasan tentang bagaimana bahasa slang berevolusi dan digunakan dalam interaksi sosial digital. Buku ini juga mengidentifikasi tren utama dalam penggunaan kata slang, seperti meningkatnya penggunaan kata slang dalam kelompok diskusi tertentu dan perubahan popularitas kata slang dari waktu ke waktu. Dengan memahami tren ini, buku memberikan kontribusi penting bagi studi linguistik digital dan analisis bahasa di platform media sosial, membuka peluang untuk analisis lebih lanjut tentang evolusi bahasa dalam ekosistem digital



**Gambar** Tren Penggunaan Kata Slang

Grafik ini menampilkan analisis tren penggunaan kata slang di media sosial dengan menggunakan teknik pengolahan bahasa alami (Natural Language Processing/NLP). Hasil dari analisis tersebut menunjukkan bahwa penggunaan kata slang mengalami peningkatan pada bulan April dan Mei, kemudian mengalami penurunan pada bulan Juni dan Juli, dan kembali meningkat pada bulan Agustus. Persentase penggunaan kata slang bervariasi antara 0,90% dan

1,00%. Analisis ini berguna untuk mengidentifikasi pola serta tren dalam penggunaan kata slang di media sosial, yang selanjutnya dapat digunakan untuk merumuskan strategi pemasaran yang lebih efektif.

### **C. Evaluasi Model NLP**

Buku ini menganalisis penggunaan kata slang di media sosial dengan menerapkan teknik pengolahan bahasa alami (Natural Language Processing atau NLP). Evaluasi terhadap model NLP yang digunakan dalam buku ini melibatkan beberapa algoritma pembelajaran mesin, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Ketiga model tersebut dievaluasi berdasarkan kinerja mereka dalam mengklasifikasikan kata slang dan tren topik yang berkembang di media sosial. Hasil evaluasi dari masing-masing model menunjukkan perbedaan dalam tingkat akurasi, presisi, recall, dan f1-score dalam mengidentifikasi kata slang dan tren yang sedang populer di platform media sosial tersebut.

Model Naïve Bayes merupakan salah satu algoritma klasifikasi yang terkenal karena

kesederhanaan dan efektivitasnya dalam pengklasifikasian teks. Pada buku ini, Naïve Bayes menunjukkan kinerja yang cukup memuaskan dalam mengidentifikasi kata slang di media sosial. Hasil evaluasi menunjukkan bahwa model ini memiliki akurasi sebesar 82%, dengan precision mencapai 80%, recall sebesar 78%, dan F1-score sebesar 79%. Model Naïve Bayes mampu menangani keragaman dalam penggunaan kata slang, meskipun memiliki keterbatasan dalam menangani hubungan kompleks antar kata. Keterbatasan ini disebabkan oleh asumsi independensi antar fitur, yang menjadi dasar dari algoritma Naïve Bayes, di mana setiap fitur dianggap tidak saling bergantung satu sama lain. Meskipun demikian, model ini tetap memberikan landasan yang kuat untuk analisis lebih mendalam terhadap penggunaan kata slang di media sosial.

Support Vector Machine (SVM) dikenal sebagai salah satu algoritma yang efektif dalam klasifikasi teks, terutama ketika data yang digunakan memiliki dimensi tinggi. Pada buku ini, SVM menunjukkan kinerja yang lebih baik dibandingkan dengan Naïve Bayes. Hasil

evaluasi menunjukkan bahwa akurasi model SVM mencapai 88%, dengan precision sebesar 85%, recall sebesar 84%, dan F1-score sebesar 84%. SVM mampu memisahkan kelas dengan lebih efektif karena memanfaatkan margin maksimal antara kelas yang berbeda. Selain itu, SVM juga mampu menangani kasus-kasus yang tidak linear dengan menggunakan kernel trick. Kelebihan ini membuat model SVM lebih handal dalam menangani variabilitas dan kompleksitas dari penggunaan kata slang di media sosial. Namun, salah satu kelemahan dari SVM adalah kebutuhan komputasi yang lebih tinggi, terutama ketika berhadapan dengan dataset yang besar.

Model Random Forest, yang merupakan kumpulan dari beberapa pohon keputusan, juga dievaluasi dalam buku ini. Random Forest dikenal karena kemampuannya dalam mengurangi overfitting dan memberikan hasil yang lebih stabil. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki akurasi sebesar 86%, dengan precision sebesar 83%, recall sebesar 81%, dan F1-score sebesar 82%. Model ini mampu menangani dataset yang tidak seimbang dengan

lebih baik karena menggunakan voting dari beberapa pohon keputusan. Random Forest juga memberikan interpretabilitas yang lebih baik dibandingkan dengan SVM, karena setiap pohon dalam hutan dapat dianalisis secara individu. Namun, meskipun akurasi model Random Forest lebih rendah dibandingkan dengan SVM, model ini menunjukkan kinerja yang konsisten dan dapat diandalkan dalam berbagai kondisi.

Dari hasil evaluasi model di atas, dapat disimpulkan bahwa model SVM menunjukkan kinerja terbaik dalam mengklasifikasikan kata slang dan trending topik di media sosial, diikuti oleh Random Forest dan Naïve Bayes. Meskipun model SVM memiliki akurasi tertinggi, Random Forest memberikan keseimbangan antara akurasi dan interpretabilitas. Sedangkan model Naïve Bayes, meskipun paling sederhana, tetap memberikan dasar yang kuat untuk analisis awal. Pemilihan model yang tepat bergantung pada kebutuhan spesifik dari buku dan sumber daya yang tersedia. Evaluasi ini memberikan gambaran yang menyeluruh tentang kinerja masing-masing model dalam analisis kata slang di media sosial, dan dapat

menjadi panduan dalam pengembangan lebih lanjut dari sistem pengolahan bahasa alami.

Dalam buku ini, kami melakukan perbandingan terhadap beberapa model pembelajaran mesin untuk menganalisis topik yang sedang tren dalam penggunaan kata slang di media sosial dengan menerapkan teknik Natural Language Processing (NLP). Model yang digunakan dalam analisis ini meliputi Naive Bayes, Support Vector Machine (SVM), dan Random Forest. Setiap model diuji menggunakan dataset yang terdiri dari 1105 tweet yang memuat kata slang. Untuk mengevaluasi performa dari masing-masing model, kami mengacu pada sejumlah metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score. Metrik-metrik ini digunakan untuk memberikan gambaran yang menyeluruh tentang kemampuan model dalam mengklasifikasikan kata slang serta mengidentifikasi tren yang muncul di media sosial..

Tabel Hasil Perbandingan 3 Metode

<b>Model</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F1-Score</b>
Naive	80%	78%	79%	78.5%
Bayes				
SVM	85%	84%	83%	83.5%
Random	75%	73%	74%	73.5%
Forest				

Tabel tersebut menampilkan hasil perbandingan tiga metode yang digunakan dalam buku ini. Berdasarkan hasil evaluasi, model Support Vector Machine (SVM) menunjukkan performa terbaik dengan akurasi sebesar 85%, presisi 84%, recall 83%, dan F1-score 83,5%. Model Naive Bayes, meskipun tidak sebaik SVM, tetap menunjukkan performa yang cukup baik dengan akurasi 80%, presisi 78%, recall 79%, dan F1-score 78,5%. Di sisi lain, model Random Forest mencatatkan performa yang lebih rendah dibandingkan kedua model lainnya, dengan akurasi 75%, presisi 73%, recall 74%, dan F1-score 73,5%. Dari hasil perbandingan ini, dapat disimpulkan bahwa model SVM lebih unggul dalam tugas klasifikasi kata slang di media sosial. Keunggulan SVM ini

kemungkinan disebabkan oleh kemampuannya dalam menangani data berdimensi tinggi serta kemampuannya memisahkan kelas-kelas dengan margin yang lebih jelas, sehingga menghasilkan klasifikasi yang lebih akurat.

Temuan dari buku ini menunjukkan bahwa model SVM mampu menangkap karakteristik unik dari kata slang di media sosial dengan lebih baik dibandingkan model lainnya. Kata slang sering kali memiliki variasi bentuk dan penggunaan yang sangat beragam, sehingga membutuhkan model yang mampu menangani kompleksitas tersebut. SVM, dengan kemampuannya menemukan hyperplane yang memisahkan data dengan margin maksimum, berhasil mengidentifikasi pola penggunaan kata slang secara lebih akurat. Selain itu, hasil buku ini juga menunjukkan bahwa pendekatan berbasis statistik seperti Naive Bayes masih relevan dan dapat memberikan hasil yang memadai, meskipun tidak seakurat SVM. Model Random Forest, meskipun mengandalkan teknik ensemble learning yang lebih kompleks, ternyata kurang efektif dalam menangani klasifikasi kata slang ini. Hal ini kemungkinan disebabkan oleh tingginya variasi penggunaan kata

slang di media sosial, yang berpotensi menyebabkan overfitting pada model Random Forest, sehingga model ini tidak mampu menangkap pola secara konsisten.

Implikasi dari temuan ini cukup signifikan untuk pengembangan lebih lanjut dalam bidang Natural Language Processing (NLP), terutama dalam analisis media sosial. Pertama, penggunaan Support Vector Machine (SVM) dapat dioptimalkan untuk tugas-tugas serupa yang melibatkan klasifikasi teks dengan karakteristik kompleks dan bervariasi, seperti analisis sentimen atau deteksi ujaran kebencian. Kemampuan SVM dalam menangani data berdimensi tinggi serta menemukan pola-pola non-linear menjadikannya pilihan yang sangat efektif dalam skenario ini. Kedua, meskipun performa model Naive Bayes tidak sebaik SVM, kesederhanaan dan efisiensi komputasionalnya tetap menjadikannya alternatif yang layak, terutama untuk aplikasi dengan keterbatasan sumber daya komputasi. Hal ini menunjukkan bahwa model yang lebih sederhana masih relevan dalam konteks tertentu, terutama ketika kecepatan dan efisiensi lebih diutamakan daripada akurasi maksimal.

Selain itu, temuan ini juga menekankan pentingnya pemilihan model yang tepat berdasarkan karakteristik data yang dihadapi. Dalam kasus kata slang di media sosial, model yang dapat menangani keragaman bentuk dan penggunaan, serta memiliki fleksibilitas dalam menemukan pola yang tidak linear, seperti SVM, cenderung memberikan hasil yang lebih baik. Implikasi lainnya adalah terbukanya peluang untuk mengeksplorasi metode kombinasi model, atau ensemble learning, yang dapat menggabungkan keunggulan dari beberapa model untuk meningkatkan akurasi dan efisiensi. Sebagai contoh, kombinasi antara SVM yang lebih akurat dengan model yang lebih cepat, seperti Naive Bayes, berpotensi menghasilkan solusi yang lebih optimal untuk aplikasi nyata dalam analisis media sosial..

Secara keseluruhan, temuan dari buku ini memberikan kontribusi yang berarti dalam pemahaman kita tentang analisis trending topik pada kata slang di media sosial dan bagaimana menggunakan teknik NLP yang tepat untuk mendapatkan hasil yang maksimal. Implementasi dari temuan ini dapat membantu

meningkatkan kualitas analisis media sosial, yang pada gilirannya dapat digunakan oleh berbagai industri untuk memahami tren dan pola komunikasi yang berkembang di platform-platform online.

Buku ini memiliki beberapa keunggulan utama. Pertama, dengan menggunakan metode Pengolahan Bahasa Alami (NLP), buku ini dapat menganalisis tren penggunaan kata slang di media sosial dengan lebih mendalam dan akurat dibandingkan metode konvensional. NLP memungkinkan pengidentifikasian pola dan kecenderungan penggunaan bahasa yang kompleks, yang sering kali tidak mudah terdeteksi dengan metode tradisional. Kedua, buku ini memanfaatkan data yang luas dan beragam dari berbagai platform media sosial, memberikan gambaran yang lebih representatif dan menyeluruh mengenai tren penggunaan kata slang. Ketiga, penggunaan algoritma machine learning memungkinkan pemrosesan data dalam jumlah besar dengan efisiensi tinggi, serta meningkatkan akurasi hasil analisis. Namun, buku ini juga memiliki beberapa keterbatasan. Salah satu keterbatasan utama adalah ketergantungan pada

kualitas dan kelengkapan data yang dikumpulkan. Data dari media sosial sering kali tidak konsisten dan dapat mengandung banyak noise, yang dapat mempengaruhi hasil analisis. Selain itu, meskipun NLP dan machine learning memiliki potensi besar, mereka masih terbatas oleh model dan algoritma yang digunakan. Algoritma yang digunakan mungkin tidak selalu mampu menangkap dan makna yang lebih dalam dari kata slang yang terus berubah. Terakhir, hasil analisis juga bergantung pada interpretasi peneliti, yang bisa mengakibatkan bias subyektif.

Buku ini memiliki implikasi praktis dan teoritis yang signifikan. Secara praktis, hasil buku ini dapat dimanfaatkan oleh berbagai pihak, termasuk pemasar, pembuat konten, dan peneliti sosial. Pemasar dapat menggunakan informasi ini untuk mengembangkan strategi pemasaran yang lebih efektif dengan memahami tren dan preferensi bahasa target audiens mereka. Pembuat konten dapat mengadopsi kata slang yang sedang tren untuk meningkatkan relevansi dan daya tarik konten mereka. Selain itu, peneliti sosial dapat menggunakan temuan ini untuk memahami dinamika

sosial dan budaya yang tercermin dalam penggunaan bahasa di media sosial.

Secara teoritis, buku ini memberikan kontribusi penting terhadap pengembangan ilmu Natural Language Processing (NLP) dan analisis bahasa. Dengan mengaplikasikan algoritma machine learning dan teknik NLP, buku ini memperluas pemahaman kita tentang dinamika perkembangan dan perubahan bahasa, khususnya dalam lingkungan digital. Selain itu, buku ini menawarkan wawasan baru tentang proses kemunculan dan penyebaran kata slang dalam masyarakat, yang tidak hanya relevan dalam konteks linguistik, tetapi juga dapat menjadi dasar untuk studi lanjutan dalam bidang sosiologi. Buku ini membuka peluang untuk mengeksplorasi lebih dalam bagaimana bahasa informal dan penggunaan kata slang berkembang di era digital, serta dampaknya terhadap interaksi sosial dan budaya dalam masyarakat yang semakin terkoneksi melalui media sosial.

Berdasarkan temuan dan keterbatasan buku ini, terdapat beberapa rekomendasi untuk buku mendatang. Pertama, diperlukan pengembangan dan

penyempurnaan algoritma NLP yang lebih canggih guna menangkap nuansa dan makna kata slang secara lebih akurat. Ini bisa dicapai dengan mengintegrasikan model deep learning yang lebih kompleks, seperti Transformer atau BERT, serta melatih model tersebut menggunakan data yang lebih beragam dan representatif dari berbagai sumber. Kedua, buku selanjutnya sebaiknya melibatkan lebih banyak platform media sosial untuk memperluas cakupan data, sehingga dapat memahami tren penggunaan kata slang di berbagai komunitas online yang berbeda. Hal ini akan memberikan gambaran yang lebih menyeluruh mengenai variasi bahasa slang di berbagai kelompok sosial.

Selain itu, penting bagi buku masa depan untuk mempertimbangkan aspek etis dalam pengumpulan dan analisis data. Penggunaan data pribadi dari media sosial harus mematuhi prinsip privasi dan mendapatkan persetujuan dari pengguna, guna menjaga integritas buku dan menghormati hak-hak individu. Terakhir, disarankan untuk mengombinasikan pendekatan kualitatif dengan kuantitatif. Pendekatan ini dapat memberikan pemahaman yang lebih mendalam tentang

konteks sosial dan budaya di balik penggunaan kata slang, misalnya dengan melakukan wawancara atau survei terhadap pengguna media sosial untuk mengungkap motivasi dan makna spesifik dari kata slang yang digunakan. Pendekatan gabungan ini akan memperkaya analisis dan memberikan wawasan yang lebih holistik mengenai fenomena bahasa slang dalam lingkungan digital.

Buku ini berhasil mengidentifikasi dan menganalisis tren penggunaan kata slang di media sosial dengan menggunakan metode Pengolahan Bahasa Alami (NLP) dan algoritma machine learning. Keunggulan utama dari buku ini adalah kemampuannya untuk mengolah data dalam jumlah besar secara efisien, sehingga dapat menghasilkan wawasan yang mendalam mengenai perkembangan tren bahasa, khususnya dalam konteks penggunaan kata slang. Analisis ini memberikan kontribusi penting terhadap pemahaman tentang dinamika bahasa yang berkembang di lingkungan digital.

Namun, buku ini juga menghadapi beberapa keterbatasan. Salah satu keterbatasan yang dihadapi

adalah kualitas data, di mana variasi dan representativitas data dapat memengaruhi hasil analisis. Selain itu, keterbatasan pada algoritma yang digunakan, meskipun efektif, masih belum mampu sepenuhnya menangkap makna yang lebih dalam dan nuansa kompleks dari kata slang. Penggunaan algoritma yang lebih canggih dan data yang lebih representatif diperlukan untuk memperbaiki hasil buku di masa mendatang.

Rekomendasi untuk buku selanjutnya mencakup pengembangan algoritma NLP yang lebih canggih, seperti penerapan model deep learning yang mampu menangkap makna dan konteks secara lebih mendalam. Selain itu, perluasan cakupan buku dengan melibatkan lebih banyak platform media sosial akan memberikan gambaran yang lebih menyeluruh tentang variasi penggunaan kata slang di berbagai komunitas online. Pendekatan yang lebih holistik juga dianjurkan, dengan menggabungkan metode kualitatif dan kuantitatif, sehingga dapat menggali motivasi dan konteks sosial di balik penggunaan kata slang.

Secara keseluruhan, buku ini memberikan kontribusi signifikan dalam bidang NLP dan analisis bahasa, serta menawarkan wawasan yang berharga bagi akademisi, profesional pemasaran, dan pihak-pihak lain yang tertarik mempelajari dinamika bahasa slang di media sosial

# DAFTAR PUSTAKA

---

- Anggraeni, M., Syafrullah, M., & Damanik, H. A. (2019). Literation Hearing Impairment (I-Chat Bot): Natural Language Processing (NLP) and Naïve Bayes Method. *Journal of Physics: Conference Series*, 1201(1), 012057. doi: 10.1088/1742-6596/1201/1/012057
- Asian, J., Rosita, M. D., & Mantoro, T. (2022). Sentiment Analysis for the Brazilian Anesthesiologist Using Multi-Layer Perceptron Classifier and Random Forest Methods. *Journal of Online Information*, 7(1), 132-141. doi: 10.15575/JOIN.V7I1.900
- Choo, S., & Kim, W. (2023). A study on the evaluation of tokenizer performance in natural language processing. *Applied Artificial Intelligence*, 37(1). doi: 10.1080/08839514.2023.2175112
- Demir, S., & Topcu, B. (2022). Graph-based Turkish text normalization and its impact on noisy text processing. *Engineering Science and Technology, an*

*International Journal*, 35, 101192. doi:  
10.1016/J.JESTCH.2022.101192

- Fatima, N., Daudpota, S. M., Kastrati, Z., Imran, A. S., Hassan, S., & Elmitwally, N. S. (2023). Improving news headline text generation quality through frequent POS-Tag patterns analysis. *Engineering Applications of Artificial Intelligence*, 125, 106718. doi: 10.1016/J.ENGAPPAI.2023.106718
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Analysis of social media Twitter with Case of Anti-Sentiment Analysis of social media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest. *Procedia Computer Science*, 161, 765-772. doi: 10.1016/j.procs.2019.11.181
- Han, K. X., Chien, W., Chiu, C. C., & Cheng, Y. T. (2020). Application of support vector machine (SVM) in the sentiment analysis of twitter dataset. *Applied Sciences*, 10(3). doi: 10.3390/app10031125
- Kampung Inggris. (2022). Bahasa Slang: Asal-Usul dan Contoh. Retrieved from

<https://www.kampunginggris.id/bahasa-slang-asal-usul-dan-contoh>

Kaware, P. D., & Raut, A. B. (2024). Automatic Detection of Multilingual Misogynistic Content in Social Media Data Based on Machine Learning Approach. *2nd International Conference on Integrated Circuits and Communication Systems (ICICACS 2024)*, 1-7. doi: 10.1109/ICICACS60521.2024.10499136

Kaya, Y. B., & Tantug, A. C. (2024). Effect of tokenization granularity for Turkish large language models. *Intelligent Systems with Applications*, 21, 200335. doi: 10.1016/J.ISWA.2024.200335

Korniichuk, R., & Boryczka, M. (2021). Averaging and boosting methods in ensemble-based classifiers for text readability. *Procedia Computer Science*, 192, 3677-3685. doi: 10.1016/j.procs.2021.09.141

Kumaresan, C., & Thangaraju, P. (2023). ELSA: Ensemble learning based sentiment analysis for diversified text. *Measurement and Sensors*, 25, 100663. doi: 10.1016/J.MEASEN.2022.100663

Luo, X. (2021). Efficient English text classification using selected Machine Learning Techniques. *Alexandria*

- Engineering Journal*, 60(3), 3401-3409. doi: 10.1016/j.aej.2021.02.009
- Madatov, K., Bekchanov, S., & Vičić, J. (2022). Dataset of stopwords extracted from Uzbek texts. *Data in Brief*, 43, 108351. doi: 10.1016/J.DIB.2022.108351
- Müller, M., Longard, L., & Metternich, J. (2022). Comparison of preprocessing approaches for text data in digital shop floor management systems. *Procedia CIRP*, 107, 179-184. doi: 10.1016/J.PROCIR.2022.04.030
- Munasaroh, S. A. (2021). Budaya Komunikasi Pada Pengguna Media Sosial Facebook Di Era New Media. *Al-Ittishol: Jurnal Komunikasi Dan Penyiaran Islam*, 2(2), 82-96.
- Nutu, M. (2021). Deep Learning Approach for Automatic Romanian Lemmatization. *Procedia Computer Science*, 192, 49-58. doi: 10.1016/J.PROCS.2021.08.006
- Nurwanda, N., Suarna, N., & Prihartono, W. (2024). PENERAPAN NLP (NATURAL LANGUAGE PROCESSING) DALAM ANALISIS SENTIMEN PENGGUNA TELEGRAM DI PLAYSTORE. *JATI*
- 62 *Monograf Analisis Sentimen Publik dan Tren Topik Pendekatan Natural Language Processing dan Machine Learning Dalam Era Digital*

(*Jurnal Mahasiswa Teknik Informatika*, 8(2), 1841-1846.

Pichiyan, V., Muthulingam, S., Sathar, G., Nalajala, S., Ch, A., & Das, M. N. (2023). Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis. *Procedia Computer Science*, 230, 193-202. doi: 10.1016/J.PROCS.2023.12.074

Rahab, H., Zitouni, A., & Djoudi, M. (2021). SANA: Sentiment analysis on newspapers comments in Algeria. *Journal of King Saud University - Computer and Information Sciences*, 33(7), 899-907. doi: 10.1016/J.JKSUCI.2019.04.012

Rothe, M., Lath, R., Kumar, D., Yadav, P., & Aylani, A. (2023). Slang language Detection and Identification In Text. *2023 14th International Conference on Computing, Communication and Networking Technologies (ICCCNT 2023)*, 1-5. doi: 10.1109/ICCCNT56998.2023.10308036

Samanta, B., Shil, R., Pal, A. R., & Pal, A. (2024). Slang Word Detection in the Context of User Profiling in the Social Media Platforms. *2024 4th International*

- Conference on Advances in Electrical, Computer, Communication and Sustainable Technologies (ICAECT 2024)*, 1-5. doi: 10.1109/ICAECT60202.2024.10468875
- Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, 121, 102342. doi: 10.1016/J.IS.2023.102342
- Urva, G., Pratiwi, M., & Syarief, A. O. (2022). Optimasi Media Sosial sebagai Pendukung Pemasaran Digital. *ABDINE: Jurnal Pengabdian Masyarakat*, 2(1), 56-61.
- Wang, J., Sun, L., Liu, Y., Shao, M., & Zheng, Z. (2022). Multimodal Sarcasm Target Identification in Tweets. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 1, 8164-8175. doi: 10.18653/V1/2022.ACL-LONG.562.

## PROFIL PENULIS

---



Shabrina Rasyid Munthe, S.Pd.I., MS, lahir di Rantauprapat pada 20 Februari 1992. Beliau menyelesaikan pendidikan S1 dalam bidang Tarbiyah di Institut Agama Islam Negeri Padangsidempuan dan meraih gelar S2 Magister Sastra dari Universitas Islam Sumatera Utara (UISU). Sejak tahun 2019 hingga saat ini, menjabat sebagai dosen tetap pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer (FILKOM), Universitas Al Washliyah Labuhanbatu.



Sudi Suryadi, S. Kom., M. Kom., lahir di Sawahlunto pada 28 Februari 1979. Beliau meraih gelar Sarjana Komputer pada Jurusan Sistem Komputer Universitas Putra Indonesia "YPTK" Padang tahun 2003, dan melanjutkan pendidikan Magister di universitas yang sama. Saat ini,

beliau aktif sebagai dosen tetap di Universitas Labuhanbatu, Fakultas Sains dan Teknologi Program Studi Sistem Informasi, serta menulis artikel dan buku referensi dalam bidang keilmuan.



Dr. Basyarul Ulya, SH., MM., lahir di Janji pada 26 April 1986, menyelesaikan studi S1 di Fakultas Hukum Universitas Islam Sumatera Utara (UISU) dari tahun 2000 hingga 2008. Beliau kemudian melanjutkan pendidikan Magister di bidang Manajemen di Universitas Islam Jakarta (UIJ) pada tahun 2009 hingga 2012, dan meraih gelar Doktor di bidang Ekonomi dari Universitas Islam Negeri Sumatera Utara (UINSU). Sejak 2017, beliau menjabat sebagai Rektor Universitas Al Washliyah Labuhanbatu.



Muhammad Rusli, MS., lahir di Pematang Siantar pada 28 April 1981. Beliau meraih gelar Sarjana Sastra pada tahun 2007 di Universitas Islam Sumatera Utara, Medan, dengan program studi Sastra Inggris. Gelar Magister Sastra diperoleh pada tahun 2012 di universitas yang sama dengan program studi Sastra Inggris. Saat ini, beliau menjabat sebagai dosen tetap di Universitas Al Washliyah Labuhanbatu.



Nurhanna Harahap, M.Hum., lahir di Teluk Rampah pada 26 Agustus 1986. Beliau meraih gelar Sarjana Pendidikan pada tahun 2009 di Universitas Muslim Nusantara Al Washliyah, Medan, dengan program studi Pendidikan Bahasa Inggris. Gelar Magister Humaniora diperoleh pada tahun 2015 di Universitas Negeri Medan, Medan, dengan program studi Linguistik Terapan Bahasa Inggris. Saat ini, beliau menjabat sebagai

dosen tetap di Universitas Al Washliyah Labuhanbatu.



Muhammad Fauzi Romadhon Marpaung, S.Pd.I, M.Hum., lahir di Aek Nabara pada 28 Maret 1992. Beliau menyelesaikan pendidikan Strata 1 di Program Studi Pendidikan Bahasa Inggris di IAIN Sumatera Utara. Pada tahun 2018, beliau meraih gelar Magister Humaniora pada Program Studi Linguistik Terapan Bahasa Inggris di Program Pascasarjana Universitas Negeri. Saat ini, beliau merupakan dosen tetap di Program Studi Pendidikan Bahasa Inggris, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Al Washliyah Labuhanbatu, Sumatera Utara, dengan karya ilmiah yang diterbitkan di jurnal nasional dan internasional.



Raden Sri Ayu Ramadhana, S.Pd .I, M.Pd lahir di Kwala Begunit pada 15 Februari 1992. Beliau menyelesaikan pendidikan S1 di Jurusan Pendidikan Matematika, Institut Agama Islam Negeri Sumatera Utara (IAIN SU), dan melanjutkan pendidikan S2 di Jurusan Pendidikan Matematika di Universitas Negeri Medan. Dari tahun 2014 hingga 2019, beliau menjadi tenaga pengajar di sekolah menengah dan Madrasah Aliyah Negeri di Langkat. Sejak tahun 2019 hingga saat ini, beliau menjabat sebagai dosen tetap pada Program Studi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan (FKIP), Universitas Al Washliyah Labuhanbatu.



Ibnu Rasyid Munthe, ST., M. Kom., lahir di Rantau Prapat, Kabupaten Labuhanbatu, pada tahun 1987. Beliau menyelesaikan pendidikan Strata 1 (S1) di Universitas Nurtanio Bandung dan melanjutkan Strata 2 (S2) di Universitas Putra Indonesia (YPTK) Padang. Aktif menulis artikel ilmiah serta buku. Saat ini, beliau mengajar di Universitas Labuhanbatu. Selain itu, beliau mengelola kanal YouTube @manjaddawajada2022 sebagai media informasi pembelajaran dan berbagi pengetahuan.



# MONOGRAF

Analisis Sentimen Publik  
dan Tren Topik Pendekatan  
*Natural Language  
Processing* dan *Machine  
Learning* dalam Era Digital

Buku ini hadir sebagai wujud dari kebutuhan yang semakin meningkat akan pemahaman mendalam terkait analisis sentimen publik dan tren topik di era digital. Seiring dengan perkembangan teknologi dan meluasnya penggunaan media sosial, kemampuan untuk memahami opini dan sentimen masyarakat menjadi sangat penting bagi berbagai sektor, termasuk bisnis, politik, hingga sosial kemasyarakatan. Buku ini menyajikan pendekatan yang komprehensif dan sistematis dalam menganalisis sentimen dan tren topik melalui teknik-teknik *Natural Language Processing (NLP)* dan *Machine Learning (ML)*. Kedua pendekatan ini memungkinkan kita untuk mengolah dan memahami data dalam skala besar secara otomatis, yang sebelumnya memerlukan banyak waktu dan tenaga jika dilakukan secara manual. Kami berharap, buku ini dapat memberikan panduan praktis bagi para peneliti, akademisi, praktisi, dan siapa saja yang tertarik dalam memahami dinamika opini publik serta bagaimana tren terbentuk dan berkembang di dunia digital. Dalam buku ini, pembaca akan diajak untuk memahami dasar-dasar teori yang mendasari *NLP* dan *machine learning*, serta bagaimana kedua teknologi ini diterapkan dalam analisis sentimen publik dan tren topik. Contoh-contoh nyata serta implementasi di berbagai sektor juga disajikan agar dapat memberikan gambaran yang lebih jelas dan aplikatif.



Penerbit  
PT ARR RAD PRATAMA  
Gunung Jati Cirebon Jawa Barat  
Indonesia 45151  
email : [arradpratama@gmail.com](mailto:arradpratama@gmail.com)

ISBN 978-623-8760-17-6



9 786238 760176