PENERAPAN *LEXICON BASED-NAÏVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP KESEHATAN MENTAL GENERASI Z

SKRIPSI

Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika

Oleh
AIDA SAFITRI
NIM. 2111017220004

PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2025

PENERAPAN *LEXICON BASED-NAÏVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP KESEHATAN MENTAL GENERASI Z

SKRIPSI

Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika

Oleh
AIDA SAFITRI
NIM. 2111017220004

PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2025

SKRIPSI

PENERAPAN *LEXICON BASED-NAIVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP KESEHATAN MENTAL GENERASI Z

Oleh:
Aida Safitri
2111017220004

Telah dipertahankan pada hari Kamis, tanggal 26 Juni 2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I

Prof. Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

Penguji I

Yuana Sukmawaty, S.Si., M.Si
NIP. 198810152015042002

Pembimbing II

Oni Soesanto, S.Si., M.Si
NIP. 197301262005011003

Penguji II

Selvi Annisa, S.Si., M.Si
NIP. 199212262022032016

Banjarbaru, 2 Juli 2025
Mengetahui,
Koordinator Program Studi Statistika
FMIPA ULM

Prof. Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 23 Juni 2025

Aida Safitri

NIM. 2111017220004

ABSTRAK

Penerapan *Lexicon Based-Naive Bayes Classifier* untuk Analisis Sentimen Publik terhadap Kesehatan Mental Generasi Z (Oleh: Aida Safitri; Pembimbing: Dewi Anggraini dan Oni Soesanto, 2025; 95 halaman)

Generasi Z adalah generasi yang lahir antara tahun 1997-2012 dan tumbuh dalam era digitalisasi sehingga teknologi menjadi salah satu bagian penting dari Generasi Z. Penggunaan teknologi yang berlebihan dapat menimbulkan dampak negatif, seperti kecanduan yang berpotensi menyebabkan ketidakstabilan emosi serta risiko masalah kesehatan mental, termasuk depresi dan kecemasan. Masalah kesehatan mental Generasi Z saat ini menjadi isu cukup serius dan banyak dibicarakan di media sosial X. Opini publik mengenai isu ini penting untuk dianalisis karena mencerminkan persepsi, pemahaman, dan sikap masyarakat terhadap kesehatan mental. Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X terhadap kesehatan mental Generasi Z menggunakan metode *Lexicon Based*, serta membandingkan kinerja model *Bernoulli* dan *Multinomial Naïve Bayes* dalam mengklasifikasikan sentimen. Hasil penelitian menunjukkan 80.6% pengguna media sosial X cenderung memberikan sentimen negatif, menggambarkan kekhawatiran publik terhadap kondisi kesehatan mental Generasi Z yang dinilai kurang sehat dan berdampak pada performa dalam kehidupan sehari-hari. Dari perbandingan kinerja, model *Multinomial Naive Bayes* menunjukkan kinerja lebih baik dibandingkan model *Bernoulli Naive Bayes* dengan sensitivitas sebesar 0.64318442, spesifisitas sebesar 0.81078431, dan *G-Mean* sebesar 0.71558742. Hasil penelitian ini, diharapkan membantu lembaga pendidikan dan perusahaan dalam merumuskan kebijakan serta program yang efektif, sekaligus meningkatkan kesadaran Generasi Z akan pentingnya menjaga kesehatan mental.

Kata Kunci: Kesehatan Mental, Generasi Z, Analisis Sentimen, *Lexicon Based*, *Naive Bayes Classifier*

ABSTRACT

Implementation of Lexicon Based-Naive Bayes Classifier for Public Sentiment Analysis on Generation Z's Mental Health (By: Aida Safitri; Advisors: Dewi Anggraini and Oni Soesanto, 2023; 95 page)

Generation Z is a generation born between 1997-2012 and grew up in an era of digitalization making technology an important part of Generation Z. Excessive use of technology has the potential to cause negative impacts, such as addiction that causes emotional instability and the risk of mental health problems, including depression and anxiety. Generation Z's mental health problems are currently a serious issue and are widely discussed on social media X. Public opinion on this issue is important to analyze as it reflects the public's perception, understanding, and attitude toward mental health. This research aims to analyze the sentiment of users on social media X towards Generation Z's mental health using a Lexicon Based method and comparing the performance of Bernoulli and Multinomial Naïve Bayes models in classifying sentiment. The results show that 80.6% of users on social media X tend to give negative sentiments, showing the public's concern about the mental health condition of Generation Z, which is considered unhealthy and has an impact on performance in daily life. From the performance comparison, the Multinomial Naive Bayes model shows better performance than the Bernoulli Naive Bayes model with a sensitivity of 0.64318442, specificity of 0.81078431, and G-Mean of 0.71558742. The results of this research are expected to help educational institutions and companies formulate effective policies and programs, while increasing Generation Z's awareness of the importance of maintaining mental health.

Keywords: Mental Health, Generation Z, Sentiment Analysis, Lexicon Based, Naive Bayes Classifier

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "Penerapan *Lexicon Based-Naive Bayes Classifier* untuk Analisis Sentimen Publik terhadap Kesehatan Mental Generasi Z". Penyusunan Tugas Akhir ini bertujuan untuk memenuhi salah satu syarat dalam menyelesaikan program sarjana S-1 di Program Studi Statistika FMIPA ULM.

Dalam proses penyusunan Tugas Akhir ini, penulis mendapatkan berbagai bantuan, bimbingan, dukungan, dan perhatian dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Orang tua dan keluarga yang selalu memberikan doa, dukungan, dan motivasi yang tiada henti selama masa perkuliahan hingga penyelesaian Tugas Akhir.
2. Ibu Prof. Dewi Anggraini, S.Si., M.App.Sci., Ph.D dan Bapak Oni Soesanto, S.Si., M.Si selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan dukungan selama proses penyusunan Tugas Akhir.
3. Ibu Yuana Sukmawaty, S.Si., M.Si dan Ibu Selvi Annisa, S.Si., M.Si selaku dosen penguji yang telah memberikan masukan dan saran selama proses penyusunan Tugas Akhir.
4. Bapak/Ibu dosen pengajar dan staf Program Studi Statistika FMIPA ULM yang telah memberikan ilmu, motivasi, dan dukungan selama masa perkuliahan.
5. Sahabat dan teman-teman yang selalu menemani, mendukung, memberikan semangat, serta mendengarkan keluh kesah penulis selama masa perkuliahan hingga penyusunan Tugas Akhir, khususnya Rizky Saputra B, Cindy Ayudia Irwanto, "Sobat Bendungan", dan "Penjoki Handal".
6. Teman-teman Program Studi Statistika angkatan 2021 yang telah berjuang bersama selama masa perkuliahan.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna dan masih terdapat kekurangan baik dari segi penulisan maupun hasil, sehingga kritik dan saran yang membangun dari semua pihak sangat diharapkan dalam membantu untuk penyempurnaan Tugas Akhir ini. Akhir kata, penulis berharap Tugas Akhir ini dapat memberikan manfaat untuk semua pihak.

Banjarbaru, 23 Juni 2025

Aida Safitri

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

DAFTAR LAMBANG DAN SINGKATAN

c	Variabel target
d	Data dengan kelas yang belum diketahui
$P(c\|d)$	Peluang dari c dengan d diketahui, disebut juga *posterior*
$P(d\|c)$	Peluang dari d dengan c diketahui, disebut juga *likelihood*
$P(c)$	Peluang c, disebut juga *prior*
$P(d)$	Peluang d, disebut juga *evidence*
t	*Term* (kata)
t_k	*Term* ke-k
V	Himpunan kata dalam data *training*
n_d	Jumlah total kata pada data d
e_i	Fitur ke-i
M	Jumlah total fitur
$P(t_k\|c)$	Peluang dari kata (*term*) t_k terdapat dalam data kelas c
N_c	Jumlah data dalam kelas c pada data *training*
N	Jumlah total data pada data *training*
T_{ct}	Jumlah kemunculan kata t dalam data *training* pada kelas c
B	Jumlah kata (*term*) unik pada semua kelas
M	Banyaknya fitur
$P(t_i\|c)$	Peluang dari kata (*term*) t_i terdapat dalam data kelas c
N_{ct}	Banyaknya data yang memuat *term* t dalam data kelas c pada data *training*
$W_{t,d}$	Bobot kata atau *term* (t) terhadap data (d)
$tf_{t,d}$	Jumlah kemunculan kata atau *term* (t) dalam data (d)
D	Jumlah total data
df_t	Jumlah data yang mengandung kata (t)
$\propto$	Proporsional
NBC	*Naïve Bayes Classifier*
BNB	*Bernoulli Naive Baye*
MNB	*Multinomial Naive Bayes*
NLP	*Natural language Processing*
ADHD	*Attention Deficit Hyperactivity Disorder*
PTSD	*Post-Traumatic Stress Disorder*
API	*Application Programming Interface*
URL	*Uniform Resource Locator*
TP	*True Positive*
TN	*True Negative*
FP	*False Positive*
FN	*False Negative*