PENERAPAN *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN PADA TEKS KESEHATAN MENTAL DENGAN METODE PENYEIMBANGAN DATA *OVERSAMPLING*

SKRIPSI

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh

GHISCA NUR ARIFA

NIM 2111016220022

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

MEI 2026

SKRIPSI

PENERAPAN *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN PADA TEKS KESEHATAN MENTAL DENGAN METODE PENYEIMBANGAN DATA *OVERSAMPLING*

Oleh:

GHISCA NUR ARIFA
2111016220022

telah dipertahankan di depan Dosen Penguji pada tanggal 26 Mei 2026

Susunan Penguji:

Pembimbing Utama	Penguji 1.
M. Itqan Mazdadi, S.Kom, M.Kom.	**Dwi Kartini, S.Kom, M.Kom.**
NIP. 199006122019031013	**NIP. 198704212012122003**
Pembimbing Pendamping	2.
Irwan Budiman, S.T., M.Kom.	**Friska Abadi, S.Kom, M.Kom.**
NIP. 197703252008121001	**NIP. 198809132023211010**

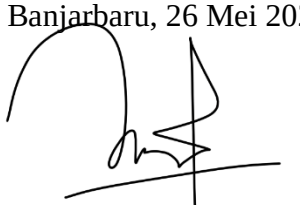
Banjarbaru, 14 Juli 2026
Koordinator Program Studi Ilmu Komputer

Dwi Kartini, S.Kom, M.Kom.
NIP. 198704212012122003

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, 26 Mei 2026



Ghisca Nur Arifa
NIM 2111016220022

ABSTRAK

PENERAPAN *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN PADA TEKS KESEHATAN MENTAL DENGAN METODE PENYEIMBANGAN DATA *OVERSAMPLING* (Oleh: Ghisca Nur Arifa; Pembimbing: Muhammad Itqan Mazdadi, S.Kom, M.Kom. dan Irwan Budiman, S.T., M.Kom.; 2026; 105 halaman)

Penelitian eksperimental ini mengkaji komparasi performa antara algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) dalam mengklasifikasikan sentimen pada teks kesehatan mental yang mengalami kendala ketidakseimbangan kelas (*imbalanced data*). Dengan memanfaatkan dataset tekstual sebanyak 53.043 baris, pengujian dilakukan berbasis pembobotan TF-IDF melalui simulasi variasi rasio pembagian data (70:30, 80:20, dan 90:10) serta penerapan empat variasi teknik oversampling, meliputi *Synthetic Minority Over-sampling Technique* (SMOTE), *Borderline*-SMOTE, dan *RandomOverSampler*. Temuan eksperimen pada dataset awal mengonfirmasi superioritas SVM berkernel linear yang mencapai akurasi tertinggi sebesar 74% karena ketangguhannya dalam memetakan batas keputusan pada ruang fitur berdimensi tinggi, sementara *Naïve Bayes* hanya mencatatkan akurasi maksimal 66%. Meskipun introduksi metode *oversampling* memicu fenomena *Accuracy Paradox*—yang ditandai dengan koreksi penurunan akurasi global menjadi 67%–72% untuk SVM dan 56%–61% untuk *Naïve Bayes* akibat risiko *noise* sintetis—intervensi ini terbukti secara signifikan mendongkrak nilai sensitivitas (*recall*) dalam mendeteksi kelas minoritas yang ekstrem seperti P*ersonality Disorder* (2,26%). Dengan demikian, pendekatan oversampling sangat direkomendasikan apabila orientasi utama sistem ditujukan untuk mengoptimalkan akurasi identifikasi pada kategori-kategori minoritas.

Kata kunci: Analisis Sentimen, Kesehatan Mental, *Naïve Bayes*, *Support Vector Machine*, *Oversampling*.

ABSTRACT

APPLICATION OF NAÏVE BAYES AND SUPPORT VECTOR MACHINE IN SENTIMENT ANALYSIS ON MENTAL HEALTH TEXTS USING OVERSAMPLING DATA BALANCING METHOD *(By: Ghisca Nur Arifa; Supervisors: Muhammad Itqan Mazdadi, S.Kom, M.Kom. and Irwan Budiman, S.T., M.Kom.; 2026; 105 pages)*

This experimental study investigates the performance comparison between Support Vector Machine (SVM) and Naïve Bayes (NB) algorithms in classifying sentiments from mental health text data characterized by severe class imbalance. Utilizing a textual dataset of 53,043 rows, the evaluation was conducted using TF-IDF weighting across three data-splitting ratios (70:30, 80:20, and 90:10) combined with four oversampling techniques, including Synthetic Minority Over-sampling Technique (SMOTE), Borderline-SMOTE, and RandomOverSampler. Experimental findings on the baseline dataset confirm the superiority of the linear-kernel SVM, which achieved a peak accuracy of 74% due to its robustness in defining decision boundaries within high-dimensional feature spaces, whereas Naïve Bayes only reached a maximum accuracy of 66%. Although the introduction of oversampling methodologies triggered the Accuracy Paradox phenomenon—characterized by a slight decline in global accuracy to 67%–72% for SVM and 56%–61% for Naïve Bayes caused by synthetic noise—this intervention significantly boosted the sensitivity (recall) in detecting extreme minority classes such as Personality Disorder (2.26%). Consequently, the oversampling approach is highly recommended when the system's primary orientation is aimed at optimizing identification accuracy within minority categories.

Keywords: Sentiment Analysis, Mental Health, Social Media, Naïve Bayes, Support Vector Machine, Oversampling

PRAKATA

Assalamu'alaikum Warrahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul "Penerapan *Naïve Bayes* dan *Support Vector Machine* dalam Analisis Sentimen pada Teks Kesehatan Mental dengan Metode Penyeimbangan Data *Oversampling*" untuk memenuhi syarat dalam penyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat serta salam pada Nabi Muhammad SAW beserta para sahabat, keluarga dan pengikut beliau hingga *yaumul qiyamah*.

Pada lembar ini penulis ingin mengucapkan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga yang selalu mendukung dalam proses penyelesaian skripsi dan menguatkan batin di saat kehilangan arah, yaitu Alm. Ayah, Ibu, Kakak Difo, Kakak Fadel, Kakak Novi dan Affan.
2. Bapak Muhammad Itqan Mazdadi, S.Kom, M.Kom. dan Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama dan pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, dan Bapak Triando Hamonangan Saragih, S.Kom, M.Kom. selaku dosen pembimbing akademik yang telah memberikan izin sehingga pengerjaan skripsi ini dapat terselesaikan.
4. Seluruh Dosen dan Staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bemanfaat.
5. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2021 yang selalu memberikan dukungan dan mengingatkan serta mendoakan dalam proses mengerjakan skripsi

6. Ucapan terima kasih kepada Kak Felia, Sethos, Nari, Hayi, Collei, Onti Furina, Tao, Mico, Ayaka, Scar, Mbaknyu, Retatang, Kak Sho dan Lili yang sudah menemani penulis dalam suka serta duka di masa pengerjaan skripsi
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 26 Mei 2026

Ghisca Nur Arifa

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN