

**KOMPARASI METODE *NAÏVE BAYES CLASSIFIER*
DAN *K-NEAREST NEIGHBOR* PADA ANALISIS
SENTIMEN ISU SAMPAH DI BALI**



TESIS

oleh

**NI PUTU ARY RARA ISWARI
NIM 2029101001**

**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA
2025**

**KOMPARASI METODE *NAÏVE BAYES CLASSIFIER*
DAN *K-NEAREST NEIGHBOR* PADA ANALISIS
SENTIMEN ISU SAMPAH DI BALI**

TESIS

Diajukan kepada
Universitas Pendidikan Ganesha
untuk Memenuhi Sebagian Persyaratan
Memperoleh Gelar Magister Komputer
Program Studi Ilmu Komputer

oleh
NI PUTU ARY RARA ISWARI
NIM 2029101001

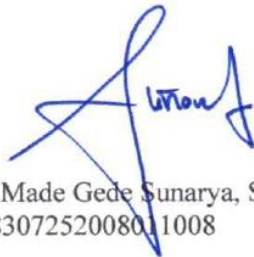
**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA
2025**

LEMBAR PERSETUJUAN PEMBIMBING

Tesis oleh Ni Putu Ary Rara Iswari ini telah diperiksa dan disetujui oleh Tim Pembimbing.

Singaraja, 14 Agustus 2025

Pembimbing I

A handwritten signature in blue ink, appearing to read 'I Made Gede Sunarya', with a large, stylized initial 'I'.

Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.
NIP 198307252008011008

Pembimbing II

A handwritten signature in blue ink, appearing to read 'I Gede Aris Gunadi', with a large, stylized initial 'I'.

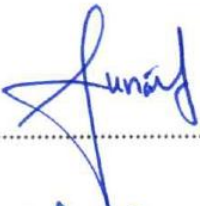
Dr. I Gede Aris Gunadi, S.Si., M.Kom.
NIP 197703182008121004

LEMBAR PERSETUJUAN PENGUJI

Tesis oleh Ni Putu Ary Rara Iswari ini telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal: 14 Agustus 2025

Oleh
Tim Penguji


....., Ketua (Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs)
NIP 198307252008011008


....., Anggota (Dr. I Gede Aris Gunadi, S.Si., M.Kom.)
NIP 197703182008121004


....., Anggota (Dr. Gede Suweken, M.Sc.)
NIP 196111111987021001


....., Anggota (Dr. Luh Joni Erawati Dewi, S.T., M.Pd.)
NIP 197606252001122001

Mengetahui Direktur
Program Pascasarjana Undiksha,

Direktur, Nyoman Jampel, M.Pd.
NIP 195910101986031003

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagianbagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 14 Agustus 2025

Yang memberi pernyataan,



Ni Putu Ary Rara Iswari

PRAKATA

Puji dan syukur penulis panjatkan ke hadapan Tuhan Yang Maha Esa atas anugerah-Nya, sehingga tesis yang berjudul “Komparasi Metode *Naïve Bayes Classifier* Dan *K-Nearest Neighbor* Pada Analisis Sentimen Isu Sampah Di Bali” dapat terselesaikan sesuai dengan yang direncanakan.

Selama penyusunan tesis ini, penulis menemui beberapa kendala dan kesulitan, namun berkat dukungan dan bantuan dari berbagai pihak, penulis dapat mengatasi hambatan dan kesulitan tersebut, sehingga pada kesempatan ini penulis ingin menyampaikan rasa terimakasih yang sebesar-besarnya kepada:

1. Bapak Rektor Universitas Pendidikan Ganesha, yang telah memberikan dukungan secara moril dan memfasilitasi berbagai kepentingan akademik, selama Penulis menempuh studi di Program Pascasarjana Universitas Pendidikan Ganesha.
2. Bapak Direktur Program Pascasarjana Universitas Pendidikan Ganesha, yang telah banyak memberikan bimbingan dan arahan serta petunjuk hingga terselesaikannya tesis ini.
3. Koordinator Program Studi Ilmu Komputer Universitas Pendidikan Ganesha yang telah banyak membantu dan menyemangati penulis selama menjalani pendidikan dan penyusunan tesis ini.
4. Bapak Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. sebagai Dosen Pembimbing I yang telah memberikan bimbingan, arahan dan petunjuk yang bermanfaat serta memacu semangat Penulis untuk menyelesaikan tesis ini sesuai yang diharapkan.
5. Bapak Dr. I Gede Aris Gunadi, S.Si., M.Kom. selaku Dosen Pembimbing II yang dengan sabar memberikan bimbingan, arahan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu mengatasi berbagai kendala dalam perjalanan studi dan penyelesaian tesis ini.
6. Bapak dan Ibu dosen dan Staf, yang telah memberikan pelajaran kepada Penulis selama mengikuti studi di Program Studi Ilmu Komputer Universitas Pendidikan Ganesha.
7. Segenap keluarga tercinta, Bapak, Ibu, dan Saudara yang telah banyak membantu memberikan semangat dan motivasi serta dukungan moril dan materiil dalam penulisan tesis ini.
8. Semua rekan, sahabat dan pihak-pihak yang tidak dapat Penulis sebutkan satu-per satu atas bantuan secara langsung maupun tidak langsung dalam penulisan tesis ini.

Akhir kata Penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penulisan tesis ini dan Penulis berharap semoga tesis ini dapat bermanfaat bagi kita semua dan dunia pendidikan di masa yang akan datang.

Singaraja, 14 Agustus 2025
Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PERSETUJUAN PENGUJI	iii
LEMBAR PERNYATAAN	iv
PRAKATA	v
DAFTAR ISI	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
ABSTRAK	xi
ABSTRACT	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Pembatasan Masalah	4
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
BAB II LANDASAN TEORI	7
2.1 Kajian Pustaka	7
2.1.1 Analisis Sentimen	7
2.1.2 <i>Preprocessing</i>	8
2.1.3 TF-IDF	9
2.1.4 <i>Naive Bayes Classifier</i> (NBC)	10
2.1.5 <i>K-Nearest Neighbor</i> (K-NN)	11
2.1.6 <i>Confusion Matrix</i>	13
2.2 Hasil Penelitian Yang Relevan	15
2.3 Kerangka Berpikir	19
BAB III METODE PENELITIAN	20
3.1 Alur Penelitian	20
3.2 Studi Pustaka	20
3.3 Pengumpulan Data	21
3.4 Pelabelan Data	22
3.5 <i>Preprocessing</i> Data	23
3.6 Proses Pembobotan Kata (TF-IDF)	26
3.7 Implementasi Klasifikasi	30

3.7.1	<i>Naïve Bayes Classifier</i>	30
3.7.2	<i>K-Nearest Neighbor</i>	34
3.8	Perbandingan Akurasi Hasil	35
BAB IV HASIL DAN PEMBAHASAN		36
4.1	Hasil Pengumpulan Data	36
4.2	Hasil Data <i>Preprocessing</i>	37
4.3	Hasil Pembobotan Data (TF-IDF)	40
4.4	Skenario Uji Coba	41
4.5	Hasil Klasifikasi	42
4.5.1	Hasil Klasifikasi <i>Naïve Bayes Classifier</i> (NBC)	42
4.5.2	Hasil Klasifikasi <i>K-Nearest Neighbor</i> (K-NN)	44
4.6	Perbandingan Hasil Metode <i>Naïve Bayes Classifier</i> (NBC) dan <i>K-Nearest Neighbor</i> (K-NN)	46
4.6.1	<i>Accuracy, Precision, Recall, dan F1-Score</i>	46
4.6.2	<i>Receiver Operating Characteristic – Area Under the Curve</i> (ROC-AUC)	49
BAB V PENUTUP		52
5.1	Rangkuman	52
5.2	Kesimpulan	53
5.3	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN		59

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi K-Nearest Neighbor	12
Gambar 3.1. Alur penelitian.....	20
Gambar 4.1 Data Awal Tweet.....	37
Gambar 4.2 Hasil Proses Normalisasi dan Cleansing.....	38
Gambar 4.3 Hasil Proses Case Folding.....	38
Gambar 4.4 Hasil Proses Stopword Removal	38
Gambar 4.5 Hasil Proses Tokenizing.....	39
Gambar 4.6 Hasil Proses Stemming.....	39
Gambar 4.7 Word Cloud dan Word Frequency Tweet Positif.....	40
Gambar 4.8 Word Cloud dan Word Frequency Tweet Negatif	40
Gambar 4.9 Hasil Pembobotan Kata (TF – IDF)	41
Gambar 4.10 Grafik Perbandingan Accuracy, Precision, Recall, dan F1-Score (Data Tidak Seimbang)	47
Gambar 4.11 Grafik Perbandingan Accuracy, Precision, Recall, dan F1-Score (Data Seimbang).....	48
Gambar 4.12 Kurva ROC NBC dan K-NN (Data Tidak Seimbang)	49
Gambar 4.13 Kurva ROC NBC dan K-NN (Data Seimbang)	50
Gambar 4.14 Grafik Perbandingan ROC-AUC	50

DAFTAR TABEL

Tabel 3.1. Tweet Isu Sampah di Bali	21
Tabel 3.2. Tweet dan Label dari Isu Sampah di Bali	22
Tabel 3.3. Hasil Proses Normalisasi	23
Tabel 3.4. Hasil Proses Cleansing.....	23
Tabel 3.5. Hasil Proses Case Folding.....	24
Tabel 3.6. Hasil Proses Tokenizing.....	24
Tabel 3.7. Hasil Proses Stopword Removal.....	25
Tabel 3.8. Hasil Proses Stemming	26
Tabel 3.9. Sampel Dokumen dan Hasil Preprocessing	26
Tabel 3.10. Proses Perhitungan TF	27
Tabel 3.11. Proses Perhitungan IDF	28
Tabel 3.12. Proses Perhitungan TF-IDF	29
Tabel 3.13. Data Latih dan Data Uji	30
Tabel 3.14. Probabilitas Setiap Kelas	31
Tabel 3.15. Probabilitas Setiap Kata Terhadap Setiap Kelas.....	32
Tabel 3.16. Perhitungan likelihood.....	33
Tabel 3.17. Perhitungan Euclidean Distance	34
Tabel 3.18. Pengurutan Nilai Euclidean Distance	34
Tabel 4.1. Contoh Tweet Positif dan Negatif.....	37
Tabel 4.2 Sub-iterasi Pengujian Metode K-NN	42
Tabel 4.3 Hasil Pengujian Naïve Bayes Classifier (NBC).....	43
Tabel 4.4 Hasil Pengujian K-Nearest Neighbor (K-NN)	45

DAFTAR LAMPIRAN

<i>Source Code Preprocessing Data</i>	59
<i>Source Code Pembobotan TF-IDF</i>	60
<i>Source Code Implementasi NB</i>	60
<i>Source Code Implementasi K-NN</i>	61

