

**ANALISIS KARAKTERISTIK *RESAMPLING* SMOTE,
SMOTE-ENN, DAN *BORDERLINE*-SMOTE
BERDASARKAN NILAI *SILHOUETTE COEFFICIENT***



PROGRAM STUDI S2 ILMU KOMPUTER

PROGRAM PASCASARJANA

UNIVERSITAS PENDIDIKAN GANESHA

2024

**ANALISIS KARAKTERISTIK *RESAMPLING* SMOTE,
SMOTE-ENN, DAN *BORDERLINE*-SMOTE
BERDASARKAN NILAI *SILHOUETTE COEFFICIENT***

TESIS

Diajukan kepada

Universitas Pendidikan Ganesha

Untuk Memenuhi Persyaratan

Memperoleh Gelar Magister Ilmu Komputer

Program Studi S2 Ilmu Komputer



Oleh

Dede Ardiansyah

NIM 2229101054

PROGRAM STUDI S2 ILMU KOMPUTER

PROGRAM PASCASARJANA

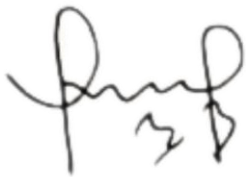
UNIVERSITAS PENDIDIKAN GANESHA

2024

Tesis oleh Dede Ardiansyah ini telah diperiksa dan disetujui untuk Mengikuti Ujian Kelayakan Tesis

Singaraja, 4 November 2024

Pembimbing I



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

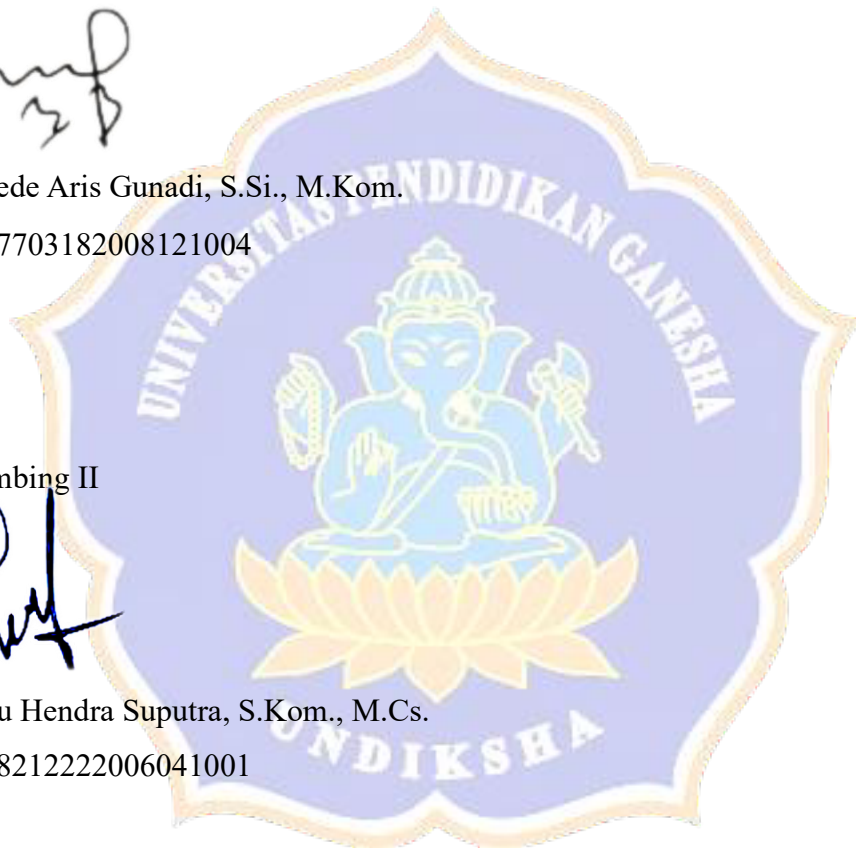
NIP 197703182008121004

Pembimbing II



Dr. Putu Hendra Suputra, S.Kom., M.Cs.

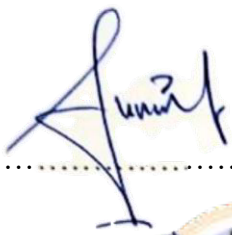

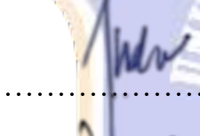
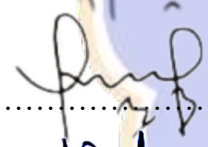

NIP 198212222006041001



Tesis oleh Dede Ardiansyah ini telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Pendidikan di Program Studi S2 Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui Pada Tanggal 30 Desember 2024

Oleh
Tim Penguji

,	Ketua (Dr. I Made Gede Sunarya, S.Kom., M.Cs.) NIP 198307252008011008
,	Anggota (Kadek Yota Ernanda, S.Kom., M.T., Ph.D.) NIP 197803242005011001
,	Anggota (Dr. Gede Indrawan, S.T., M.T.) NIP 197601022003121001
,	Anggota (Dr. I Gede Aris Gunadi, S.Si., M.Kom.) NIP 197703182008121004
,	Anggota (Dr. Putu Hendra Suputra, S.Kom., M.Cs) NIP 198212222006041001

Mengetahui Direktur
Program Pascasarjana Undiksha,



Prof. Dr. I Nyoman Jampel, M.Pd.
NIP 195910101986031003

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Pendidikan 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 bagian-bagian 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,
Yang memberi pernyataan

Dede Ardiansyah

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas anugerah-Nya, sehingga tesis yang berjudul: “Analisis Karakteristik *Resampling* Smote, Smote-Enn, dan *Borderline*-Smote Berdasarkan Nilai *Silhouette Coefficient*”, dapat diselesaikan sesuai dengan yang direncanakan.

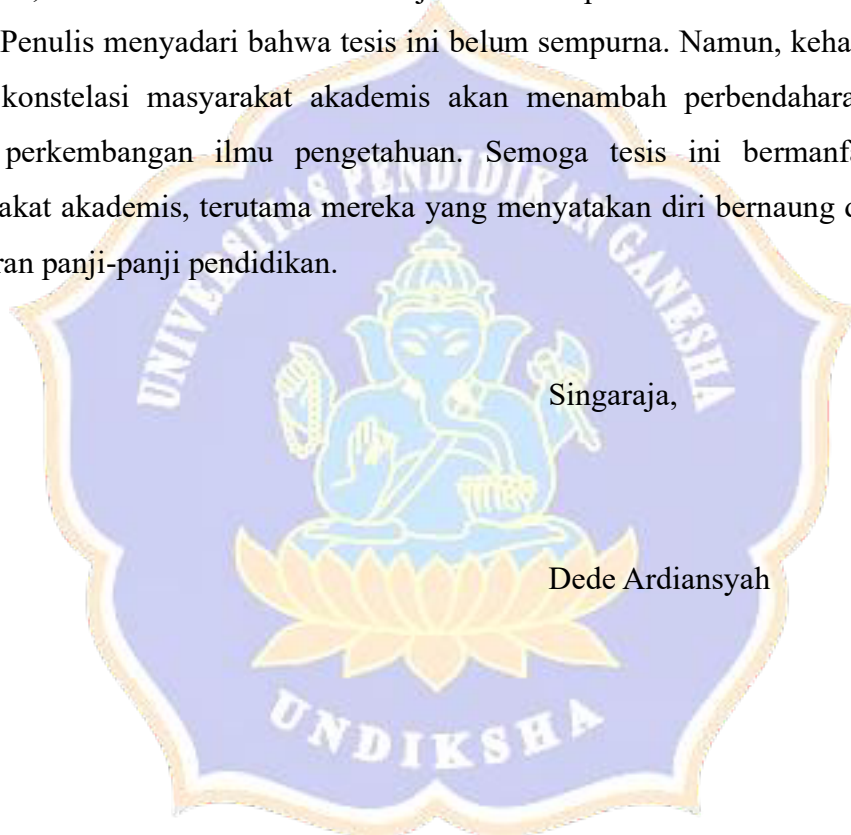
Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Pendidikan Pascasarjana Universitas Pendidikan Ganesha pada Program Studi S2 Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, izinkan penulis menyampaikan terima kasih dan penghargaan kepada pihak-pihak berikut.

1. Dr. I Gede Aris Gunadi, S.Si., M.Kom., sebagai pembimbing I yang telah dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai hambatan dalam perjalanan studi dan penyelesaian tesis ini;
2. Dr. Putu Hendra Suputra, S.Kom., M.Cs., sebagai pembimbing II, yang dengan gaya dan pola komunikasi yang khas, telah melecut semangat, motivasi, dan harapan penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis ini dapat terwujud dengan baik sesuai harapan;
3. Kadek Yota Ernanda, S.Kom., M.T., Ph.D. dan Dr. Gede Indrawan, S.T., M.T., sebagai penguji yang telah banyak memberikan masukan-masukan yang bermanfaat untuk penyempurnaan tesis ini;
4. Koordinator Program Studi Dr. I Made Gede Sunarya, S.Kom., M.Cs. dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini;
5. Direktur Program Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis menyelesaikan tesis ini;
6. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moral dan memfasilitasi berbagai kepentingan penulis dalam menyelesaikan tesis ini;

7. Rekan-rekan seangkatan di Program Studi S2 Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kedirian penulis selama menjalani studi dan penyelesaian tesis ini;
8. Bapak Suprpto dan Ibu Yeni Wirawati selaku orang tua penulis, yang telah banyak membantu secara material dan moral selama penyelesaian tesis ini.

Semoga semua bantuan yang telah mereka berikan dalam menyelesaikan studi ini, mereka diberkati imbalan yang sepadan oleh Tuhan Yang Maha Esa, kesehatan, dan keharmonian dalam menjalani kehidupan.

Penulis menyadari bahwa tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah perbendaharaan ilmu dalam perkembangan ilmu pengetahuan. Semoga tesis ini bermanfaat bagi masyarakat akademis, terutama mereka yang menyatakan diri bernaung di bawah kebesaran panji-panji pendidikan.



Singaraja,

Dede Ardiansyah

DAFTAR ISI

LEMBAR PERNYATAAN	iii
PRAKATA.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN.....	13
1.1 Latar Belakang Masalah	13
1.2 Identifikasi Masalah	14
1.3 Pembatasan Masalah	15
1.4 Rumusan Masalah	16
1.5 Tujuan Penelitian.....	16
1.6 Manfaat Penelitian.....	16
BAB II KAJIAN PUSTAKA	18
2.1 Ketidakseimbangan data.....	18
2.2 Resampling.....	19
2.3 SMOTE.....	21
2.4 SMOTE ENN	28
2.5 <i>Borderline</i> SMOTE	32
2.6 Silhouette Coefficient.....	37
2.7 Kajian Penelitian yang Relevan.....	43
BAB III METODE PENELITIAN	51

3.1 Jenis Penelitian	51
3.2 Prosedur Penelitian	51
3.2.1 Preprocessing Data	52
3.2.2 <i>Resampling</i> Data	54
3.2.3 Silhouette Coefficient	57
3.2.4 Penarikan Kesimpulan	58
BAB IV PENUTUP	59
4.1 Preprocessing Data	59
4.1.1 Deskripsi data	59
4.1.2 Pembentukan Kondisi Dataset Berdasarkan <i>Silhouette Coefficient</i>	60
4.2 Hasil <i>Resampling</i>	60
4.2.1 SMOTE	61
4.2.2 SMOTE-ENN	65
4.2.3 Borderline-SMOTE	68
4.4 Pembahasan	74
BAB V PENUTUP	81
5.1 Kesimpulan	81
5.2 Saran	82
DAFTAR PUSTAKA	84

DAFTAR TABEL

Tabel 2.1 Data Perhitungan SMOTE	22
Tabel 2.2 Kelas Minoritas Perhitungan SMOTE	24
Tabel 2.3 Jarak Euclidean Proses SMOTE	24
Tabel 2.4 Hasil Sampel Sintetis Pertama pada Proses SMOTE	26
Tabel 2.5 Hasil Proses SMOTE	26
Tabel 2.6 Data Hasil SMOTE	28
Tabel 2.7 Pemilihan Jarak Terdekat SMOTE-ENN	29
Tabel 2.8 Hasil Proses SMOTE-ENN	30
Tabel 2.9 Contoh Data Untuk Borderline-SMOTE	32
Tabel 2.10 Jarak Euclidean Untuk Proses Borderline-SMOTE	33
Tabel 2.11 Sampel Sintetis Pertama Borderline-SMOTE	34
Tabel 2.12 Hasil Proses Borderline-SMOTE	35
Tabel 2.13 koefisien Silhouette	38
Tabel 2.14 Data untuk Silhouette Coefficient Data 1	38
Tabel 2.15 Jarak Titik dengan Jarak Euclidean Data 1	39
Tabel 2.16 Perhitungan a_i Data 1	39
Tabel 2.17 Data untuk Silhouette Coefficient Data 2	40
Tabel 2.18 Jarak Titik dengan Jarak Euclidean Data 2	40
Tabel 2.19 Nilai b_i dan a_i Data 2	41
Tabel 2.20 Hasil Silhouette Coefficient	42
Tabel 2.21 Kajian Penelitian Relevan	43
Tabel 4.1 Deskripsi Data	59
Tabel 4. 2 karakteristik data awal shilouette negatif	60
Tabel 4. 3 Jumlah outlier SMOTE	63
Tabel 4. 4 Nilai silhouette SMOTE	64
Tabel 4. 5 Jumlah outlier SMOTE-ENN	66
Tabel 4. 6 Nilai silhouette SMOTE-ENN	67
Tabel 4. 7 Jumlah outlier Borderline-SMOTE	70

Tabel 4. 8 Nilai silhouette Borderline-SMOTE	71
Tabel 4. 9 Akurasi SVM Dataset Awal.....	72
Tabel 4. 10 Akurasi SVM SMOTE	72
Tabel 4. 11 Akurasi SVM SMOTE-ENN.....	72
Tabel 4. 12 Akurasi SVM Borderline-SMOTE.....	73
Tabel 4. 13 Perbandingan Akurasi SVM sebelum dan sesudah diresampling	73
Tabel 4. 14 Analisis Korelasi	76



DAFTAR GAMBAR

Gambar 3.1 Prosedur Penelitian.....	52
Gambar 3.2 Preprocessing Data.....	53
Gambar 4.1 Silhouette <i>Index</i> Negatif.....	60
Gambar 4. 2 Proses tahapan pertama metode SMOTE.....	61
Gambar 4. 3 Proses tahapan terakhir metode SMOTE	62
Gambar 4. 4 Jumlah outlier pada data 1.....	62
Gambar 4. 5 Nilai silhouette pada data 1	63
Gambar 4. 6 Jumlah outlier SMOTE	64
Gambar 4. 7 Nilai silhouette SMOTE.....	65
Gambar 4. 8 Proses tahapan pertama SMOTE-ENN.....	65
Gambar 4. 9 Proses tahapan terakhir SMOTE-ENN	66
Gambar 4. 10 Jumlah Outlier Smote-enn.....	67
Gambar 4. 11 Nilai silhouette SMOTE-ENN	68
Gambar 4. 12 Proses tahapan pertama Borderline-SMOTE	69
Gambar 4. 13 Proses tahapan terakhir Borderline-SMOTE.....	69
Gambar 4. 14 Jumlah outlier Borderline-SMOTE	70
Gambar 4. 15 Nilai silhouette Borderline-SMOTE	71
Gambar 4. 16 Data sebelum SMOTE	75
Gambar 4. 17 Data setelah SMOTE.....	75
Gambar 4. 18 Data sebelum SMOTE-ENN	77
Gambar 4. 19 Data setelah SMOTE-ENN	77
Gambar 4. 20 Data Sebelum Borderline-SMOTE	79
Gambar 4. 21 Data Setelah Borderline-SMOTE.....	79