

**STUDI KOMPARASI METODE SUPPORT VECTOR
MACHINES DAN RANDOM FOREST UNTUK
KLASIFIKASI KATA KERJA
BAHASA JEPANG**



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

**STUDI KOMPARASI METODE SUPPORT VECTOR
MACHINES DAN RANDOM FOREST UNTUK
KLASIFIKASI KATA KERJA
BAHASA JEPANG**



TESIS

Diajukan kepada

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

oleh

I NYOMAN SWETA

NIM 2129101005

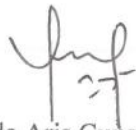
**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA
2025**

LEMBAR PERSETUJUAN PEMBIMBING

Tesis oleh I Nyoman Sweta ini telah diperiksa dan disetujui untuk mengikuti Ujian
Tesis

Singaraja, 6 Agustus 2025

Pembimbing I



Dr. I Gede Aris Gunadi, S.Si. M.Kom.
NIP. 19770318 200812 1 004

Pembimbing II



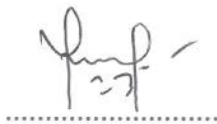
Dr. I Made Gede Sunarya, S.Kom., M.Cs.
NIP. 1983072520 0801 1 008

LEMBAR PERSETUJUAN

Tesis oleh I Nyoman Sweta 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. I Gede Aris Gunadi, S.Si., M.Kom.
NIP. 197703182008121004



Anggota Prof. Dr. I Made Candiasa, MI. Komp.
NIP. 196012311986011004



Anggota Dr. I Made Agus Wirawan, S.Kom., M.Cs.
NIP. 198408272008121001



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



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

Mengetahui
Dekan Pascasarjana Undiksha,


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

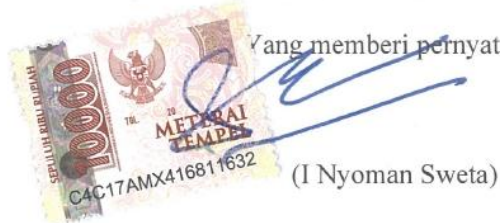
LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Pendidikan dari 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, 14 Agustus 2025

Yang memberi pernyataan,



(I Nyoman Sweta)

PRAKATA

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia Nya, sehingga penulis dapat menyelesaikan tugas akhir tesis yang berjudul “Studi Komparasi Metode Support Vector Machines dan Random Forest untuk Klasifikasi Kata Kerja Bahasa Jepang”. Penelitian ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi Ilmu Komputer, Pascasarjana Universitas Pendidikan Ganesha.

Dalam proses penyusunan tesis ini, penulis memperoleh banyak bimbingan, dukungan, dan motivasi dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setulus tulusnya kepada:

1. Dr. I Gede Aris Gunadi, S.Si. M.Kom., selaku Pembimbing I, atas bimbingan, arahan, dan motivasi yang sangat berarti selama proses penyusunan tesis ini.
2. Dr. I Made Gede Sunarya, S.Kom., M.Cs., selaku Pembimbing II, atas saran, masukan, dan evaluasi yang membantu penulis dalam menyempurnakan penelitian ini.
3. Prof. Dr. I Made Candiasa, MI. Komp., selaku Dosen Penguji I, atas masukan berharga, arahan kritis, dan pandangan konseptual yang sangat memperkaya isi dan kualitas penelitian ini.
4. Dr. I Made Agus Wirawan, S.Kom., M.Cs., selaku Dosen Penguji II, atas pertanyaan mendalam, koreksi teknis, serta saran aplikatif yang membantu penulis melihat penelitian ini dari berbagai perspektif.

5. Kepala Program Studi Ilmu Komputer beserta seluruh dosen pengajar di Pascasarjana Undiksha atas ilmu dan bimbingan yang diberikan selama masa studi.
6. Direktur Pascasarjana Universitas Pendidikan Ganesha, atas dukungan dan fasilitas yang diberikan dalam mendukung proses penyusunan tesis ini.
7. Rektor Universitas Pendidikan Ganesha, atas kesempatan yang diberikan untuk menempuh pendidikan di lingkungan akademik yang kondusif.
8. Teman teman seperjuangan di Program Studi Ilmu Komputer, atas kebersamaan, semangat, dan dukungan selama proses studi dan penelitian berlangsung.
9. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Pimpinan BPS Provinsi Bali atas dukungan, arahan, dan fasilitas yang telah diberikan selama proses penyusunan penelitian ini.
10. Ucapan terima kasih juga penulis sampaikan kepada rekan-rekan di BPS Provinsi Bali yang telah memberikan bantuan, kerja sama, serta semangat selama proses pengumpulan data dan penyusunan penelitian ini. Dukungan yang tulus dan kerja sama yang baik menjadi salah satu faktor penting dalam terselesaikannya penelitian ini.
11. Kedua orang tua dan keluarga tercinta, atas doa, kasih sayang, dan dukungan moral yang tiada henti hingga tesis ini dapat diselesaikan.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan, baik dari segi teori maupun implementasi. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan manfaat bagi pembaca, serta menjadi kontribusi kecil dalam pengembangan ilmu pengetahuan, khususnya di bidang ilmu komputer dan linguistik komputasional.

Singaraja, 14 Agustus 2025

Penulis



DAFTAR ISI

	(halaman)
LEMBAR PERSETUJUAN PEMBIMBING.....	ii
LEMBAR PERSETUJUAN TIM PENGUJI.....	iii
LEMBAR PERNYATAAN	iv
PRAKATA.....	v
ABSTRAK	viii
ABSTRACT.....	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang.....	16
1.2 Identifikasi Masalah	20
1.3 Batasan Penelitian.....	21
1.4 Rumusan Masalah	22
1.5 Tujuan Penelitian.....	23
1.6 Manfaat Penelitian.....	24
BAB II KAJIAN PUSTAKA	25
2.1 Pengertian dan Klasifikasi Kata Kerja Bahasa Jepang	25
2.1.1 Kelompok Kata Kerja Bahasa Jepang.....	25
2.1.2 Akhiran “ru” dalam Klasifikasi Kata Kerja	27
2.2 Metode Klasifikasi	27
2.2.1 <i>Support Vector Machine (SVM)</i>	28
2.2.2 <i>Random forest</i>	31
2.3 Ekstraksi Fitur	32
2.4 Kajian Hasil Penelitian yang Relevan	33
2.5 Metode Evaluasi.....	36
BAB III METODE PENELITIAN.....	40
3.1 Desain Penelitian	40
3.2 Pengumpulan Data	41
3.2.1 Sumber Data	41
3.2.2 Jenis Data yang Dikumpulkan.....	41
3.2.3 Validasi Data	42
3.3 <i>Preprocessing</i> Data.....	43
3.3.1 Pembersihan Data.....	43
3.3.2 Konversi ke Romaji.....	44

3.3.3	Penghapusan Duplikasi Data	45
3.4	Ekstraksi dan rekayasa fitur	46
3.4.1	Ekstraksi Fitur	46
3.4.2	Rekayasa Fitur	48
3.5	Pemodelan dan Implementasi Algoritma	50
3.5.1	Implementasi dengan <i>SVM</i>	51
3.5.2	Implementasi dengan <i>RF</i>	51
3.6	Evaluasi Model	52
3.6.1	Metrik Evaluasi	52
3.6.2	Perbandingan Performa Model	57
3.7	Analisis dan Interpretasi Hasil	58
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		60
4.1	Deskripsi Dataset dan Preprocessing	60
4.1.1	Karakteristik <i>Dataset</i>	61
4.1.2	Tahapan <i>Preprocessing</i>	62
4.2	Ekstraksi dan Transformasi Fitur	63
4.2.1	Deskripsi Fitur yang Digunakan	63
4.2.2	Transformasi dengan <i>One Hot Encoding</i>	64
4.3	Hasil Klasifikasi dengan <i>SVM</i>	66
4.3.1	Konfigurasi Model <i>SVM</i>	66
4.3.2	Evaluasi performa <i>SVM</i>	67
4.3.3	<i>Confusion Matrix</i> dan Interpretasi	68
4.3.4	Matrik evaluasi	70
4.3.5	<i>ROC Curve</i> dan Interpretasi	71
4.4	Hasil Klasifikasi dengan <i>Random Forest</i>	73
4.4.1	Konfigurasi model <i>Random Forest</i>	73
4.4.2	Evaluasi Performa <i>Random Forest</i>	73
4.4.3	<i>Confusion matrix</i> dan Interpretasi	75
4.4.4	Metrik Evaluasi	76
4.4.5	<i>ROC Curve</i> dan Interpretasi	77
4.5	Perbandingan Performa <i>SVM</i> dan <i>Random Forest</i>	79
4.5.1	Perbandingan Metrik Evaluasi	79
4.5.2	Kesimpulan Perbandingan	81
4.6	Analisis Kesalahan Klasifikasi	82
4.6.1	Pola Kesalahan dari <i>Confusion Matrix</i>	83
4.6.2	Pola Kesalahan Klasifikasi	83
4.6.3	Konsistensi Pola Kesalahan antar Model	85
4.7	Evaluasi Model Berdasarkan Data Latih dan Uji	86

BAB V PENUTUP.....	89
5.1 Kesimpulan.....	89
5.2 Saran.....	90
DAFTAR PUSTAKA.....	92
LAMPIRAN.....	94



DAFTAR TABEL

Tabel 1. 1 Peta Konversi Tag Kata Kerja	17
Tabel 2. 1 Kata Kerja K1	26
Tabel 2. 2 Kata Kerja K2	26
Tabel 2. 3 Kata Kerja K3	27
Tabel 2. 4 Perbandingan terhadap Penelitian-Penelitian Terkait	34
Tabel 3. 1 Data Kata Kerja Bahasa Jepang	42
Tabel 3. 2 Penghapusan Karakter yang Tidak Relevan	43
Tabel 3. 3 Konversi ke Romaji.....	44
Tabel 3. 4 Penghapusan Duplikasi Data.....	45
Tabel 3. 5 Fitur ContainRU.....	46
Tabel 3. 6 Fitur OcbRU (One Character before RU).....	47
Tabel 3. 7 Fitur TcbRU (Two Characters before RU)	48
Tabel 3. 8 <i>One Hot Encoding</i> Fitur OcbRU	49
Tabel 3. 9 <i>One Hot Encoding</i> Fitur TcbRU.....	49
Tabel 3. 10. Rekapitulasi Jumlah Fitur.....	50
Tabel 3. 11 Confusion Matrix	55
Tabel 4. 1 Ringkasan distribusi kelompok	61
Tabel 4. 2 Transformasi <i>OcbRU</i>	65
Tabel 4. 3 Evaluasi Performa <i>SVM</i>	67
Tabel 4. 4 Evaluasi Probabilistik dan Efisiensi Komputasi Model <i>SVM</i>	68
Tabel 4. 5 Evaluasi Performa <i>Random Forest (RF)</i>	74
Tabel 4. 6 Evaluasi Probabilistik dan Efisiensi Komputasi Model <i>RF</i>	74
Tabel 4. 7 Perbandingan Metrik Evaluasi <i>SVM</i> dan <i>RF</i>	80
Tabel 4. 8 Pola kesalahan dari confusion matrix	83
Tabel 4. 9 Contoh Tumpang Tindih Morfologis Kata Kerja Berakhiran -ru	84
Tabel 4. 10 Metrik Evaluasi antara Data Latih dan Uji pada <i>SVM RF</i>	86

DAFTAR GAMBAR

Gambar 3. 1 Desain Penelitian	40
Gambar 4. 1 <i>Confusion Matrix SVM</i>	69
Gambar 4. 2 Perbandingan <i>Precision, Recall, dan F1-Score SVM</i>	70
Gambar 4. 3 <i>ROC Curve</i> Multikelas Model SVM (OvR)	72
Gambar 4. 4 <i>Confusion Matrix Random Forest</i>	75
Gambar 4. 5 Perbandingan <i>Precision, Recall, dan F1-Score RF</i>	76
Gambar 4. 6 <i>ROC Curve</i> Multikelas Model RF	78
Gambar 4. 7 Perbandingan <i>Log Loss</i>	80
Gambar 4. 8 Perbandingan Akurasi dan F1-score antara Data Latih dan Uji	87



DAFTAR LAMPIRAN

LAMPIRAN 1. One Hot Encoding Fitur TcbRU.....	94
---	----

