

**KLASIFIKASI MULTIKATEGORI BERITA
OLAHRAGA INDONESIA MENGGUNAKAN
ALGORITMA KNN DAN SVM**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh

Ardan Ardiansyah

227006516138



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2026

HALAMAN PENGESAHAN
TUGAS AKHIR

KLASIFIKASI MULTIKATEGORI BERITA OLAH RAGA INDONESIA
MENGUNAKAN ALGORITMA KNN DAN SVM



Ardan Ardiansyah
227006516138

Dosen Pembimbing

(Gatot Soepriyono, S.Si., M.S.M.)

NIDN. 0317047803

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

KLASIFIKASI MULTIKATEGORI BERITA OLAHRAGA INDONESIA
MENGUNAKAN ALGORITMA KNN DAN SVM

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, 24 Februari 2026



Ardan Ardiansyah

227006516138

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

KLASIFIKASI MULTIKATEGORI BERITA OLAHRAGA INDONESIA MENGGUNAKAN ALGORITMA KNN DAN SVM

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026



Dosen Pembimbing

Gatot Soepriyono, S.Si., M.S.M.

NIDN. 0317047803

Ketua Program Studi

Dr. Ir. Andrianingsih, S.Kom., MMSI

NIDN. 0303097902

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Ardan Ardiansyah
NPM : 227006516138
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 24 Februari 2026




JUDUL DALAM BAHASA INDONESIA :

KLASIFIKASI MULTIKATEGORI BERITA OLAHRAGA INDONESIA
MENGUNAKAN ALGORITMA KNN DAN SVM

JUDUL DALAM BAHASA INGGRIS :

MULTI-CATEGORY CLASSIFICATION OF INDONESIAN SPORTS NEWS
USING KNN AND SVM ALGORITHMS

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Mahasiswa	Ko-Prodi
TGL : 02/02/2026	TGL :	TGL : 24/02/2026
		

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “**KLASIFIKASI MULTIKATEGORI BERITA OLAHRAGA INDONESIA MENGGUNAKAN ALGORITMA KNN DAN SVM**” Skripsi ini disusun sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sarjana Fakultas Teknologi Komunikasi dan Informatika.

Proses penelitian hingga penyusunan skripsi ini tentunya tidak terlepas dari dukungan, doa, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Bapak Gatot Soepriyanto, S.Si., M.S.M., selaku dosen pembimbing utama yang telah memberikan arahan, bimbingan, masukan, serta meluangkan waktu dan tenaga dalam membimbing penulis selama proses penelitian dan penyusunan skripsi ini.
2. Ibu dan Ayah tercinta yang senantiasa memberikan doa, dukungan moral maupun materi, serta motivasi yang tiada henti selama masa perkuliahan hingga terselesaikannya skripsi ini.
3. Seluruh dosen pengajar di Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika maupun dosen dari program studi lainnya yang telah memberikan ilmu, wawasan, dan pengalaman berharga selama penulis menempuh pendidikan.
4. Teman-teman seangkatan serta sahabat-sahabat terdekat Fajar, Pancar, Ando, Iqbal, Ahmad, Haikal, Alfad, Maisaan dan Naufal yang telah memberikan dukungan, semangat, bantuan, serta kebersamaan selama proses perkuliahan dan penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan demi penyempurnaan karya ini di masa yang akan datang. Semoga

skripsi ini dapat memberikan manfaat dan kontribusi, khususnya dalam pengembangan ilmu di bidang Teknologi Informasi.

Jakarta, 24 Februari 2026



Ardan Ardiansyah

ABSTRAK

Perkembangan portal berita daring di Indonesia menyebabkan meningkatnya volume berita olahraga yang dipublikasikan setiap hari, sehingga proses pengelompokan berita secara manual menjadi kurang efisien dan berpotensi menimbulkan inkonsistensi kategorisasi. Penelitian ini mengkaji dan membandingkan kinerja algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan berita olahraga berbahasa Indonesia ke dalam tiga subkategori, yaitu Sepak Bola, Badminton, dan MotoGP. Representasi teks dilakukan menggunakan metode *Term Frequency Inverse Document Frequency (TF-IDF)* setelah melalui tahapan *preprocessing teks* yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Dataset diperoleh melalui proses *Web Scraping* portal berita daring Detik.com dan dibagi ke dalam dua skenario pembagian data, yaitu 70:30 dan 80:20 menggunakan metode *stratified sampling*. Hasil eksperimen menunjukkan bahwa kedua algoritma menghasilkan performa yang sangat tinggi dengan akurasi melebihi 99% pada kedua skenario pengujian. Pada pembagian 70:30, KNN dan SVM memperlihatkan hasil yang relatif identik, sedangkan pada skenario 80:20, SVM menunjukkan performa yang sedikit lebih unggul dan lebih stabil dibandingkan KNN. Validasi menggunakan *Stratified K-Fold Cross Validation* serta analisis *macro average* dan *learning curve* mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik tanpa gejala *overfitting* yang signifikan. Implementasi sistem berbasis *Streamlit* juga membuktikan bahwa model mampu mengklasifikasikan berita olahraga aktual secara langsung. Berdasarkan hasil penelitian, algoritma SVM dinilai sebagai model yang paling optimal dalam klasifikasi multikategori berita olahraga berbahasa Indonesia.

Kata kunci: klasifikasi teks, berita olahraga, TF-IDF, K-Nearest Neighbor, Support Vector Machine.

ABSTRACT

The development of the online news portal in Indonesia has led to an increase in the volume of sports news published daily, making the manual news grouping process less efficient and potentially leading to categorization inconsistencies. This study examines and compares the performance of the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms in classifying Indonesian-language sports news into three subcategories: Football, Badminton, and MotoGP. Text representation was performed using the Term Frequency Inverse Document Frequency (TF-IDF) method after going through pre-processing stages including casefolding, tokenizing, stopword removal, and stemming. The dataset was obtained through Web Scraping of the online news portal Detik.com and divided into two data split scenarios, namely 70:30 and 80:20 using the stratified sampling method. Experimental results show that both algorithms produce very high performance with accuracy exceeding 99% in both test scenarios. In the 70:30 split, KNN and SVM displayed relatively identical results, while in the 80:20 scenario, SVM showed slightly superior and more stable performance than KNN. Validation using Stratified K-Fold Cross Validation, along with macro-average and learning curve analysis, demonstrated that the model had good generalization capabilities without significant overfitting. The implementation of the Streamlit-based system also demonstrated the model's ability to classify current sports news directly. Based on the research results, the SVM algorithm was deemed the most optimal model for multi-category classification of Indonesian-language sports news.

Keywords: *text classification, sports news, TF-IDF, K-Nearest Neighbor, Support Vector Machine.*

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iii
LEMBAR PERSETUJUAN REVIEW AKHIR	iv
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	v
KATA PENGANTAR.....	vi
abstrak	viii
abstract.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur.....	6
2.2 Justifikasi Pemilihan Algoritma K-Nearest Neighbor dan Support Vector Machine	16
2.3 Landasan Teori.....	18
2.3.1 Text Mining	18
2.3.2 Preprocessing Teks.....	18
2.3.3 <i>Case folding</i>	19

2.3.4	<i>Tokenizing</i>	19
2.3.5	<i>Stopword removal</i>	19
2.3.6	<i>Stemming</i>	20
2.3.7	TF-IDF (Term Frequency Inverse Document Frequency)	20
2.3.8	K-Nearest Neighbor (KNN).....	21
2.3.9	<i>Support Vector Machine (SVM)</i>	22
2.3.10	Evaluasi Model.....	23
BAB III METODOLOGI PENELITIAN		26
3.1	Jenis dan Pendekatan Penelitian.....	26
3.2	Waktu Penelitian.....	26
3.3	Subjek Penelitian.....	27
3.4	Fokus Penelitian	27
3.5	Sumber dan Pengumpulan Data	28
3.5.1	Sumber Data	28
3.5.2	Proses Pengambilan Data.....	29
3.5.3	Struktur Data Hasil Scraping.....	30
3.6	Tahapan Penelitian	31
3.7	<i>Preprocessing</i> Teks	32
3.7.1	<i>Case folding</i>	33
3.7.2	<i>Tokenizing</i>	33
3.7.3	<i>Stopword removal</i>	34
3.7.4	<i>Stemming</i>	34
3.8	Proses Ekstraksi Fitur Menggunakan TF-IDF.....	35
3.9	Tahapan Pelatihan dan Pengujian Model Klasifikasi	36
3.9.1	Proses Klasifikasi Menggunakan Algoritma K-Nearest Neighbor	

(KNN)	37
3.9.2 Proses Klasifikasi Menggunakan Algoritma <i>Support Vector Machine (SVM)</i>	38
3.10 Metode Evaluasi Model.....	40
3.11 Confusion Matrix	40
BAB IV HASIL DAN PEMBAHASAN	42
4.1 Hasil Data Scrapping.....	42
4.2 Hasil <i>Preprocessing teks</i> Data.....	42
4.2.1 <i>Case folding</i>	43
4.2.2 <i>Tokenizing</i>	43
4.2.3 <i>Stopword removal</i>	44
4.2.4 <i>Stemming</i>	44
4.2.5 Hasil <i>Preprocessing</i>	45
4.3 Hasil Ekstraksi Fitur TF-IDF	46
4.4 Hasil Pembagian Data	47
4.5 Perhitungan Manual Algoritma	48
4.5.1 Perhitungan Manual Algoritma K-Nearest Neighbor (KNN)	48
4.5.2 Perhitungan Manual Algoritma <i>Support Vector Machine (SVM)</i>	51
4.6 Hasil Implementasi Algoritma	54
4.6.1 Hasil Implementasi Algoritma KNN (70:30)	54
4.6.2 Hasil Implementasi Algoritma SVM (70:30)	56
4.6.3 Hasil Implementasi Algoritma KNN (80:20)	59
4.6.4 Hasil Implementasi Algoritma SVM (80:20)	61
4.7 Hasil Perbandingan Evaluasi Model	63
4.8 Validasi Model.....	65

4.8.1	Validasi Menggunakan Stratified K-Fold Cross Validation.....	65
4.8.2	Analisis Nilai <i>Macro average</i>	66
4.8.3	Analisis <i>Learning curve</i>	68
4.9	Implementasi Sistem Klasifikasi Berita	69
4.9.1	Pengujian Sistem Menggunakan Berita Aktual.....	70
BAB V KESIMPULAN DAN SARAN		72
5.1	Kesimpulan.....	72
5.2	Saran.....	72
DAFTAR PUSTAKA		74
LAMPIRAN		78



DAFTAR GAMBAR

Gambar 3. 1 Proses Pengambilan Data	29
Gambar 3. 2 Tahapan Penelitian.....	32
Gambar 3. 3 Proses TF-IDF	35
Gambar 3. 4 Proses Klasifikasi KNN.....	37
Gambar 3. 5 Proses Klasifikasi SVM.....	39
Gambar 4. 1 Hasil Scrapping Data Olahraga	42
Gambar 4. 2 Hasil <i>Case folding</i>	43
Gambar 4. 3 Hasil <i>Tokenizing</i>	44
Gambar 4. 4 Hasil <i>Stopword removal</i>	44
Gambar 4. 5 Hasil <i>Stemming</i>	45
Gambar 4. 6 Hasil Preprocessing	45
Gambar 4. 7 Hasil TF-IDF	47
Gambar 4. 8 Confusion Matrix KNN 70:30.....	54
Gambar 4. 9 Hasil Model KNN 70:30.....	56
Gambar 4. 10 Hasil Confusion Matrix SVM 70:30.....	57
Gambar 4. 11 Hasil Model SVM 70:30.....	58
Gambar 4. 12 Hasil Confusion Matrix KNN 80:20.....	59
Gambar 4. 13 Hasil Model KNN 80:20.....	61
Gambar 4. 14 Hasil Confusion Matrix SVM 80:20.....	61
Gambar 4. 15 Hasil Model SVM 80:20.....	63
Gambar 4. 16 Hasil Perbandingan Evaluasi Model	64
Gambar 4. 17 Hasil Stratified K-Fold Cross Validation	65
Gambar 4. 18 Hasil <i>Macro average</i>	67
Gambar 4. 19 Hasil <i>Learning curve</i>	68
Gambar 4. 20 Tampilan Antarmuka Sistem	69
Gambar 4. 21 Pengujian Sistem Menggunakan Berita Olahraga Aktual	

DAFTAR TABEL

Tabel 3. 1 Waktu Penelitian.....	26
Tabel 3. 2 Pembagian Data Latih	36
Tabel 4. 1 Data Latih	48
Tabel 4. 2 Data Uji	48
Tabel 4. 3 Rekapitulasi Hasil Perhitungan Manual KNN.....	50
Tabel 4. 4 Penentuan Kelas	51
Tabel 4. 5 Nilai Bobot dan Bias	52
Tabel 4. 6 Data Uji SVM.....	52
Tabel 4. 7 Rekapitulasi Hasil Perhitungan Manual SVM	53

