



SKRIPSI

**KLASIFIKASI DAN PEMETAAN JUDI ONLINE
BERDASARKAN ARTIKEL BERITA DENGAN
MENGUNAKAN ALGORITMA SVM DAN NER**

WISNU MUKTI DARWANSAH

NPM 21082010146

DOSEN PEMBIMBING

Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Rizka Hadiwiyanti, S.Kom, M.Kom, MBA.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA**



SKRIPSI

**KLASIFIKASI DAN PEMETAAN JUDI ONLINE
BERDASARKAN ARTIKEL BERITA DENGAN
MENGUNAKAN ALGORITMA SVM DAN NER**

WISNU MUKTI DARWANSAH

NPM 21082010146

DOSEN PEMBIMBING

Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Rizka Hadiwiyantri, S.Kom, M.Kom, MBA.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

KLASIFIKASI DAN PEMETAAN JUDI ONLINE BERDASARKAN ARTIKEL BERITA DENGAN MENGGUNAKAN ALGORITMA SVM DAN NER

Oleh :
WISNU MUKTI DARWANSAH
NPM. 21082010146

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi
Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal
10 Juli 2025

Menyetujui

Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 201803 2 001

..... (Pembimbing I)

Rizka Hadiwivanti, S.Kom, M.Kom, MBA
NIP. 19860727 2018032 001

..... (Pembimbing II)

Eka Dvar Wahvuni, S.Kom, M.Kom
NIP. 19841201 202121 1 005

..... (Ketua Penguji)

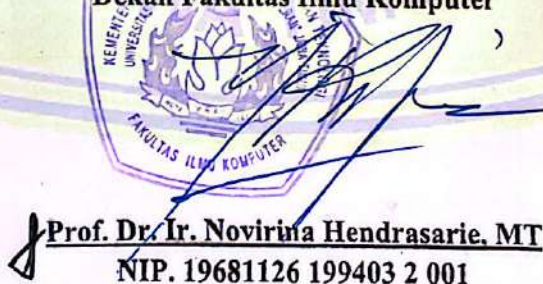
Reisa Permatasari, ST, M.Kom
NIP. 19920514 202203 2007

..... (Penguji II)

Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 212199 10 320267

..... (Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**KLASIFIKASI DAN PEMETAAN JUDI ONLINE BERDASARKAN
ARTIKEL BERITA DENGAN MENGGUNAKAN ALGORITMA SVM
DAN NER**

Oleh:

**WISNU MUKTI DARWANSAH
NPM. 21082010146**



Menyetujui,

**Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer**

Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : WISNU MUKTI DARWANSAH
NPM : 21082010146
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 23 Juli 2025
Yang Membuat Pernyataan,



WISNU MUKTI DARWANSAH
NPM. 21082010146

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Wisnu Mukti Darwansah / 21082010146
Judul Skripsi : Klasifikasi Dan Pemetaan Judi Online Berdasarkan Artikel Berita Dengan Menggunakan Algoritma Svm Dan Ner
Dosen Pembimbing : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom.
2. Rizka Hadiwiyanti, S.Kom, M.Kom, MBA.

Fenomena judi online di Indonesia telah berkembang pesat dan menjadi ancaman serius secara sosial maupun ekonomi. Skripsi ini bertujuan untuk mengklasifikasikan dan memetakan aktivitas judi online berdasarkan berita digital dengan menggunakan algoritma Support Vector Machine (SVM) dan Named Entity Recognition (NER). Data dikumpulkan dari portal berita Detik.com, Kompas.com, dan Tribunnews pada periode 2017 hingga 2024 melalui metode web scraping. Tahapan penelitian dimulai dengan setup dan import library, upload data, eksplorasi data, pelabelan data sesuai Undang-Undang No. 1 Tahun 2023, pra-pemrosesan, filtering data, normalisasi dan ekstraksi lokasi, serta pembersihan data lokasi. Selanjutnya dilakukan pelatihan model SVM untuk klasifikasi risiko, diikuti dengan prediksi. Evaluasi dilakukan menggunakan akurasi dan F1-score untuk mengukur kinerja model secara keseluruhan dan keseimbangan klasifikasi. Berdasarkan hasil evaluasi, model SVM Normal menunjukkan kinerja terbaik dengan akurasi 96,94% dan nilai F1-score sebesar 0,98. Hasil penelitian ini menunjukkan bahwa kombinasi NER dan SVM mampu mengidentifikasi lokasi dan tingkat risiko aktivitas judi online dengan baik. Diharapkan hasil penelitian ini dapat memberikan kontribusi bagi aparat penegak hukum dan pembuat kebijakan dalam upaya pencegahan dan penindakan aktivitas judi online di Indonesia.

Kata kunci: Judi Online, Support Vector Machine (SVM), Named Entity Recognition (NER), Klasifikasi Risiko, Visualisasi Data

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Wisnu Mukti Darwansah / 21082010146
Thesis Title : Classification and Mapping of Online Gambling
Based on News Articles Using SVM and NER
Algorithms
Advisor : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom.
2. Rizka Hadiwiyanti, S.Kom, M.Kom, MBA.

The phenomenon of online gambling in Indonesia has developed rapidly, posing serious social and economic threats. This thesis aims to classify and map online gambling activities based on digital news using the Support Vector Machine (SVM) algorithm and Named Entity Recognition (NER). Data were collected from the news portals Detik.com, Kompas.com, and Tribunnews from 2017 to 2024 through a web scraping approach. The research process included setup and library import, data upload, data exploration, data labeling according to Law No. 1 of 2023, data preprocessing, data filtering, location normalization and extraction, and location data cleaning. Subsequently, the SVM model was trained for risk classification and followed by prediction. Evaluation was conducted using accuracy and F1-score metrics to assess overall model performance and classification balance. Based on the evaluation results, the Normal SVM model demonstrated the best performance with an accuracy of 96.94% and an F1-score of 0.97. The findings indicate that the combination of NER and SVM effectively identifies the location and risk level of online gambling activities. This research is expected to contribute to law enforcement authorities and policymakers in their efforts to prevent and address online gambling activities in Indonesia.

Keywords: *Online Gambling, Support Vector Machine (SVM), Named Entity Recognition (NER), Risk Classification, Data Visualization.*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul “Klasifikasi Dan Pemetaan Judi Online Berdasarkan Artikel Berita Dengan Menggunakan Algoritma Svm Dan Ner” dapat terselesaikan dengan baik. Laporan skripsi ini disusun sebagai salah satu persyaratan untuk mendapatkan gelar Sarjana Ilmu Komputer pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur. Oleh karena itu, pada kesempatan ini penulis menyampaikan penghargaan dan ucapan terima kasih secara khusus kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
2. Bapak Agung Brastama Putra, S.Kom., M.Kom., selaku Koordinator Program Studi Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas segala dukungan yang telah diberikan sehingga proses penyusunan skripsi ini dapat berjalan dengan lancar.
3. Ibu Amalia Anjani Arifiyanti, S.Kom., M.Kom., selaku Dosen Pembimbing I, yang dengan penuh kesabaran telah meluangkan waktu untuk membimbing, mengarahkan, serta memberikan masukan yang sangat berarti selama proses penulisan skripsi.
4. Ibu Rizka Hadiwiyanti, S.Kom., M.Kom., MBA., selaku Dosen Pembimbing II, yang telah banyak memberikan arahan, bimbingan, serta masukan yang konstruktif. Penulis sangat berterima kasih atas kesabaran dan perhatian beliau dalam mendampingi proses bimbingan skripsi ini.
5. Seluruh dosen di Program Studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas ilmu, motivasi, dan dukungan yang diberikan selama masa studi.
6. Kedua orang tua dan seluruh keluarga tercinta, atas doa, cinta, dan pengorbanan yang tanpa henti menguatkan penulis dalam setiap langkah. Terima kasih telah menjadi tempat pulang terbaik dan sumber semangat terbesar sepanjang proses ini.
7. Seseorang yang *spesial*, yang senantiasa hadir memberikan semangat, doa, dan dukungan secara emosional di setiap fase perjuangan penulis. Terima kasih atas

kesabaran dalam menghadapi segala dinamika selama proses ini, serta atas motivasi yang tidak pernah surut bahkan di saat penulis merasa kehilangan arah. Kehadiranmu memberikan ketenangan dan menjadi penyemangat besar dalam menyelesaikan skripsi ini.

8. Rekan-rekan di Cahaya Kreatif, yang telah menjadi bagian penting dalam perjalanan penulis, baik secara profesional maupun pribadi. Terima kasih atas kerja sama yang solid, suasana kerja yang menyenangkan, serta dukungan dan pengertian selama penulis menjalani proses penyusunan skripsi ini.
9. Teman-teman Garangan, KM Langgeng, dan HimaTeras, yang telah memberikan warna dalam kehidupan perkuliahan penulis. Terima kasih atas tawa yang tulus, candaan yang menyegarkan, serta momen kebersamaan yang telah menjadi kebahagiaan di tengah kesibukan akademik.
10. Teman-teman seperjuangan skripsi, yang telah menjadi tempat berbagi cerita, saling menyemangati, berdiskusi, bahkan mengeluh di tengah tekanan penyusunan tugas akhir. Terima kasih atas kebersamaan, kerja sama, dan tawa yang membuat ruang skripsi tidak sekadar tempat kerja, tapi juga ruang bertumbuh bersama.
11. Seluruh pihak yang telah berkontribusi dan tidak dapat disebutkan satu per satu, atas segala bantuan dan dukungannya hingga skripsi ini dapat diselesaikan dengan baik.

Penulis menyadari sepenuhnya bahwa dalam penyusunan skripsi ini masih terdapat sejumlah kekurangan. Oleh karena itu, penulis sangat mengharapkan masukan dan saran konstruktif dari berbagai pihak demi perbaikan dan penyempurnaan di masa mendatang. Besar harapan penulis, karya ini dapat memberikan manfaat bagi pembaca serta menjadi rujukan yang berguna bagi penelitian selanjutnya.

Surabaya, Juni 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK.....	ix
KATA PENGANTAR	xiii
DAFTAR ISI	xv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	4
1.4. Tujuan Skripsi.....	5
1.5. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu.....	7
2.2 Landasan Teori	9
2.2.1. Portal Berita Digital	10
2.2.2. Judi Online	11
2.2.3. Text Mining	12
2.2.4. Klasifikasi	13
2.2.5. Pipeline	13
2.2.6. Eksplorasi Data Analisis.....	14
2.2.7. Direct Base Learning	15
2.2.8. Pengambilan Data	16
2.2.9. Preprocessing Data	16
2.2.9.1. Membersihkan Data Duplikat	16
2.2.9.2. Menghapus Nomer.....	17
2.2.9.3. Menghapus Tanda Baca	17
2.2.9.4. Menghapus Spasi Berlebih	17
2.2.9.5. Normalisasi Teks	17
2.2.10. Named Entity Recognition	18
2.2.11. Support Vector Machine.....	18
2.2.12. Oversampling	19
2.2.13. Pembobotan TF-IDF	20
2.2.14. ConfusionMatrix.....	21
2.2.15. Accuracy.....	21
2.2.16. Precision	22
2.2.17. Recall	22
2.2.18. F1-Score	22
2.2.19. Bahasa Pemrograman Python	23
2.2.20. Logs	23
2.2.21. Google Collab.....	24
2.2.22. Geopy	24
2.2.23. Flask	25
2.2.24. Next.js.....	25
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....	27

3.1 Metode Penelitian	27
3.1.1. Identifikasi Masalah	28
3.1.2. Tinjauan Pustaka	28
3.1.3. Analisis Kebutuhan Pembangunan Model	28
3.1.4. Pengambilan Data	29
3.1.5. Preprocessing Data	30
3.1.6. Pelabelan	32
3.1.7. Eksplorasi Data Analis	34
3.1.8. NER (Named Entity Recognizer).....	35
3.1.9. Training Model.....	38
3.1.10. Evaluasi Model.....	42
3.1.11. Uji Model	43
3.1.12. Visualisasi Data.....	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1. Hasil.....	45
4.1.1. Pengambilan Data	45
4.1.2. Preprocessing Data	51
4.1.3. Pelabelan.....	54
4.1.4. Eksplorasi Data Analis	62
4.1.5. NER (Named Entity Recognizer)	68
4.1.6. Training Model	77
4.1.7. Evaluasi Model	88
4.1.8. Uji Model.....	100
4.1.9. Visualisasi Data	104
4.2. Pembahasan	126
BAB V PENUTUP	129
5.1. Kesimpulan	129
5.2. Saran	129
DAFTAR PUSTAKA	131
LAMPIRAN.....	137

DAFTAR GAMBAR

Gambar 3.1. Diagram Alur Tahapan Skripsi	27
Gambar 3.2. Diagram Alur Pembagian Data	38
Gambar 4.1 Kode Web Scraping Judul Berita	49
Gambar 4.2 Dataset Hasil Scraping	50
Gambar 4.3 Kode Untuk Pra-proses	52
Gambar 4.4 Kode Frekwensi Kata	57
Gambar 4.5 Potongan Kode Untuk Proses Pelabelan	61
Gambar 4.6 Potongan Kode Distribusi Label	64
Gambar 4.7 Tampilan Grafik Distribusi Label	65
Gambar 4.8 Kode Frekuensi Singkatan Wilayah	67
Gambar 4.9 Tampilan Frekuensi Singkatan Wilayah	68
Gambar 4.10 Potongan Kode Normalisasi Dict Dari Sebagian Wilayah	69
Gambar 4.11 Kode Untuk Normalisai Nama Wilayah Di Indonesia	70
Gambar 4.12 Kode Ekstrasi Lokasi	74
Gambar 4.13 Potongan Kode Filter Data	76
Gambar 4.14 Kode Pembagian Data 80:20	79
Gambar 4.15 Kode Pembagian Data 70:30	80
Gambar 4.16 Kode Penerapan Model SVM Normal	81
Gambar 4.17 Kode SVM normal Kombinasi TF-IDF	83
Gambar 4.18 Kode Penerapan Oversampling	84
Gambar 4.19 Kode SVM Oversampling	86
Gambar 4.20 Kode SVM Oversampling Kombinasi TF-IDF	88
Gambar 4.21 Hasil SVM Normal	89
Gambar 4.22 Hasil SVM Normal Kombinasi TF-IDF	90
Gambar 4.23 Hasil SVM Oversampling	91
Gambar 4.24 Hasil SVM Oversampling Kombinasi TF-IDF	92
Gambar 4.25 Hasil SVM CountVectorizer 70:30	93
Gambar 4.26 Hasil SVM Kombinasi TF-IDF 70:30	94
Gambar 4.27 Hasil SVM Oversampling 70:30	95
Gambar 4.28 Hasil SVM Oversampling Kombinasi TF-IDF 70:30	96
Gambar 4.29 Ringkasan Hasil Pemodelan	98

Gambar 4.30 Kode Pengambilan Model Terbaik	99
Gambar 4.31 Kode Uji Model	102
Gambar 4.32 Hasil Uji Model.....	103
Gambar 4.33 Tampilan Halaman Unggah Dataset	104
Gambar 4.34 Tampilan Ringkasan Data Scraping.....	105
Gambar 4.35 Tampilan Proses Data	106
Gambar 4.36 Tampilan Status Proses Real-Time	107
Gambar 4.37 Tampilan Peta dan Tabel Persebaran Risiko.....	108
Gambar 4.38 Potongan Kode Frontend Upload File	109
Gambar 4.39 Potongan Kode Backend Upload File.....	111
Gambar 4.40 Potongan Kode Frontend Preproses	112
Gambar 4.41 Potongan Kode Backend Preproses	113
Gambar 4.42 Potongan Kode Frontend Data Filtering	114
Gambar 4.43 Potongan Kode Backend Data Filtering.....	115
Gambar 4.44 Potongan Kode Frontend Ekstraksi Lokasi.....	116
Gambar 4.45 Potongan Kode Backend Ekstraksi Lokasi	117
Gambar 4.46 Potongan Kode Frontend Normalisasi Lokasi	118
Gambar 4.47 Potongan Kode Backend Normalisasi Lokasi.....	119
Gambar 4.48 Potongan Kode Frontend Prediksi SVM.....	120
Gambar 4.49 Potongan Kode Backend Prediksi SVM	121
Gambar 4.50 Potongan Kode Backend Clean Lokasi.....	122
Gambar 4.51 Potongan Kode Frontend Get Clean Data.....	123
Gambar 4.52 Potongan Kode Backend Clean Lokasi.....	125

DAFTAR TABEL

Tabel 2. 1 Jumlah dan Nilai Transaksi Terkait Judi Online di Indonesia Tahun (2017-2023)	9
Tabel 2. 2 Media Online yang Paling Banyak Digunakan Warga Indonesia 2024	10
Tabel 3. 1 Tabel Entitas Spacy	37
Tabel 3. 2 Skenario Dalam Klasifikasi SVM	39
Tabel 4.1 Contoh Hasil Dari Tahap Praproses	53
Tabel 4.2 Contoh Sample Kata Berdasarkan Term Frequency	58
Tabel 4.3 Justifikasi Pelabelan Risiko Judi Online Berdasarkan UU No. 1 Tahun 2023	58
Tabel 4.4 Contoh Data Hasil Pelabelan	62
Tabel 4.2 Contoh Data Terekstraksi	75
Tabel 4.3 Perbandingan Hasil Keseluruhan Model	97

Halaman ini sengaja dikosongkan