

**PENGUNAAN NATURAL LANGUAGE
PROCESSING DENGAN ALGORITMA SUPPORT
VECTOR MACHINES UNTUK KLASIFIKASI GENRE
NOVEL BAHASA INDONESIA**

SKRIPSI

Oleh:

Dian Firstian Alfarabi

202010225027



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2024**

LEMBAR PERSETUJUAN PEMBIMBING

(DIGUNAKAN UNTUK TUGAS AKHIR)

Judul Skripsi : Penggunaan *Natural Language Processing*
dengan Algoritma *Support Vector Machines*
untuk Klasifikasi Genre Novel Bahasa Indonesia.

Nama Mahasiswa : Dian Firstian Alfarabi

Nomor Pokok Mahasiswa : 202010225027

Program Studi/Fakultas : Informatika/Ilmu Komputer

Jakarta, 02 Agustus 2024

MENGESAHKAN

Pembimbing I

Ketua Program Studi



Adi Muhajirin, S.Kom., M.Kom., M.M.
NIP: 0318038501



Ahmad Fathurozi, S.E., M.M.S.I
NIP: 2012486

Program Studi Informatika

Fakultas Ilmu Komputer

Universitas Bhayangkara Jakarta Raya

2024

LEMBAR PENGESAHAN

Judul Skripsi : Penggunaan *Natural Language Processing*
dengan Algoritma *Support Vector Machines*
untuk Klasifikasi Genre Novel Bahasa
Indonesia.

Nama Mahasiswa : Dian Firstian Alfarabi

Nomor Pokok Mahasiswa : 202010225027

Program Studi / Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 25 Juli 2024

Jakarta, 02 Agustus 2024

MENGESAHKAN

Ketua Tim Penguji : Wowon Priatna, S.T., M.T.I
NIDN : 0429118007

Penguji I : Dian Hartanti, S.Kom., M.M.S.I
NIDN : 0329098303

Penguji II : Adi Muhajirin, S.Kom., M.Kom., M.M.
NIDN : 0318038501

MENGETAHUI,

Ketua
Program Studi Informatika

Ahmad Fathurrozi, S.E., M.M.S.I
NIP. 2012486

Dekan
Fakultas Ilmu Komputer

Dr. Dra. Tyastuti Sri Lestari, M.M.
NIP. 1408206



**UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA**

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Dian Firstian Alfarabi
NPM : 202010225027
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Penggunaan *Natural Language Processing* dengan
Algoritma *Support Vector Machines* untuk Klasifikasi
Genre Novel Bahasa Indonesia.

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan hasil karya saya sendiri dan benar keasliannya. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 02 Agustus 2024
Penulis



Dian Firstian Alfarabi

ABSTRAK

Dian Firstian Alfarabi . 202010225027 . Penggunaan Natural Language Processing dengan Algoritma Support Vector Machines untuk Klasifikasi Genre Novel Bahasa Indonesia. Bekasi: Fakultas Ilmu Komputer . Universitas Bhayangkara Jakarta Raya. 2023

Novel adalah salah satu seni karya tulis yang sangat populer dan diminati banyak masyarakat, karena biasanya memiliki gaya penulisan yang mudah dipahami. Wattpad adalah salah satu platform terbesar dalam penyedia novel gratis yang dapat diakses dari seluruh dunia. Oleh karena itu, klasifikasi pada isi novel untuk menentukan isi teks novel tersebut agar dapat dikategorikan ke dalam genre yang ditentukan sangatlah penting. Hal ini mempermudah para penulis dan pembaca untuk menentukan genre pada novel yang mereka tulis maupun baca. Namun platform wattpad ini memiliki algoritma yang hanya menampilkan novel-novel dengan popularitas tinggi dari daftar genre, sehingga popularitas novel yang tidak memiliki genre hanya dan hanya bergantung terhadap hashtag saja kurang populer. Penelitian ini menggunakan metode Natural Language Processing (NLP) untuk melakukan pra-pemrosesan terhadap novel berbahasa Indonesia dan algoritma Support Vector Machines (SVM) untuk memungkinkan keberhasilan klasifikasi yang akurat. Pra-pemrosesan mencakup tokenisasi, stemming, dan transformasi menggunakan TF-IDF. Model SVM yang dilatih dengan data latih menghasilkan akurasi sebesar 92% dalam menentukan isi dari novel. Uji coba pada sembilan novel menunjukkan bahwa enam novel memiliki label positif dengan genre romantis, sedangkan tiga novel lainnya memiliki label negatif, menunjukkan bahwa novel tersebut tidak mengandung genre romantis. Hasil ini menunjukkan bahwa metode yang digunakan efektif dalam mengklasifikasikan genre novel, dan dapat membantu penulis dan pembaca dalam mengidentifikasi genre novel dengan lebih mudah dan akurat. Penelitian ini juga memberikan saran untuk pengembangan lebih lanjut, termasuk perluasan dataset, eksplorasi metode lain, dan peningkatan teknik pra-pemrosesan.

Kata kunci: Novel, Natural Language Processing (NLP), Support Vector Machines (SVM), Klasifikasi Genre, Pra-pemrosesan, Wattpad, Bahasa Indonesia.

ABSTRACT

Dian Firstian Alfarabi. 202010225027. *The Utilization of Natural Language Processing with Support Vector Machines Algorithm for Classification Genres in Indonesian Novels.* Bekasi: Faculty of Computer Science. Bhayangkara University Jakarta Raya. 2023.

Novels are a popular form of written art that many people enjoy due to their generally easy-to-understand writing style. Wattpad is one of the largest platforms providing free novels accessible worldwide. Therefore, classifying the content of novels to categorize them into specific genres is very important. This helps writers and readers determine the genre of the novels they write or read. However, Wattpad's platform features an algorithm that only displays novels with high popularity from the genre list, making novels without a specific genre less popular since their popularity relies solely on hashtags. This research uses Natural Language Processing (NLP) methods for preprocessing Indonesian-language novels and the Support Vector Machines (SVM) algorithm to achieve accurate classification. Preprocessing includes tokenization, stemming, and transformation using TF-IDF. The SVM model trained with training data achieved an accuracy of 92% in determining the content of novels. Tests on nine novels showed that six novels were positively labeled as romantic, while the other three were negatively labeled, indicating that those novels did not contain the romantic genre. These results demonstrate that the method used is effective in classifying novel genres and can assist writers and readers in identifying novel genres more easily and accurately. This research also provides suggestions for further development, including expanding the dataset, exploring other methods, and improving preprocessing techniques.

Keywords: Novel, Natural Language Processing (NLP), Support Vector Machines (SVM), Genre Classification, Pre-processing, Wattpad, Indonesian Language

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Dian Firstian Alfarabi
NPM : 202010225027
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif** (*Non-Exclusive Royalty-Free Right*), atas karya ilmiah saya yang berjudul :

Penggunaan *Natural Language Processing* dengan Algoritma *Support Vector Machines* untuk Klasifikasi Genre Novel Bahasa Indonesia.

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi

Pada tanggal : 02 Agustus 2024

Yang Menyatakan



KATA PENGANTAR

Segala puji bagi Allah SWT, yang telah melimpahkan rahmat dan hidayah- Nya sehingga saya dapat menyelesaikan Penelitian Skripsi yang berjudul **“Penggunaan Natural Language Processing dengan Algoritma Support Vector Machines untuk Klasifikasi Genre Novel Bahasa Indonesia”** yang saya teliti di Universitas Bhayangkara Jakarta Raya selama 3 bulan lamanya. Namun masih ada kewajiban yang harus peneliti kerjakan yaitu pembuatan laporan tertulis Penelitian Skripsi. Dengan adanya Penelitian Skripsi, saya mendapatkan banyak sekali wawasan, dan ilmu yang bermanfaat. Skripsi ini dikerjakan sebagai tanggung jawab dan bukti tertulis akan syarat lulus pada mata kuliah skripsi yang saya ambil pada semester 8 ini.

Dalam penyusunan proposal penelitian skripsi ini saya banyak sekali mendapatkan bantuan dari berbagai pihak, oleh sebab itu saya ingin mengungkapkan rasa terima kasih kepada :

1. Bapak Irjen Pol (Purn) Dr. Drs. Bambang Karsono, S.H., M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Bapak Adi Muhajirin, S.Kom., M.Kom., M.M. selaku Dosen Pembimbing I dalam penyusunan penelitian skripsi Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
5. Bapak Ajif Yunizar Pratama Yusuf, S.Si., M.Eng selaku Dosen Pembimbing Akademik.
6. Para Ibu dan Bapak Dosen beserta Seluruh Staff Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
7. Kepada Orang Tua saya yang telah memberikan dukungan moral maupun material.
8. Rekan-rekan seperjuangan fasilkom Angkatan 2020 yang telah memberikan dorongan dan semangat dalam penyusunan penelitian skripsi.
9. *I extend my deepest gratitude to myself for the unwavering commitment, and tireless efforts in the composition of this skripsi. It's a testament to my*

perseverance from inception to completion which has been navigated with determination and self-belief.

Penelitian skripsi yang saya susun ini, mungkin masih terdapat banyak kekurangan dan masih jauh dari kata sempurna, oleh karena itu saya dengan hati terbuka menerima masukan baik berupa kritik, maupun saran-saran yang dapat membangun untuk memperbaiki kekurangan yang ada. Semoga skripsi ini bermanfaat, baik bagi saya pribadi maupun bagi orang lain yang membacanya..

Bekasi, 02 Agustus 2024

Hormat saya



Dian Firstian Alfarabi



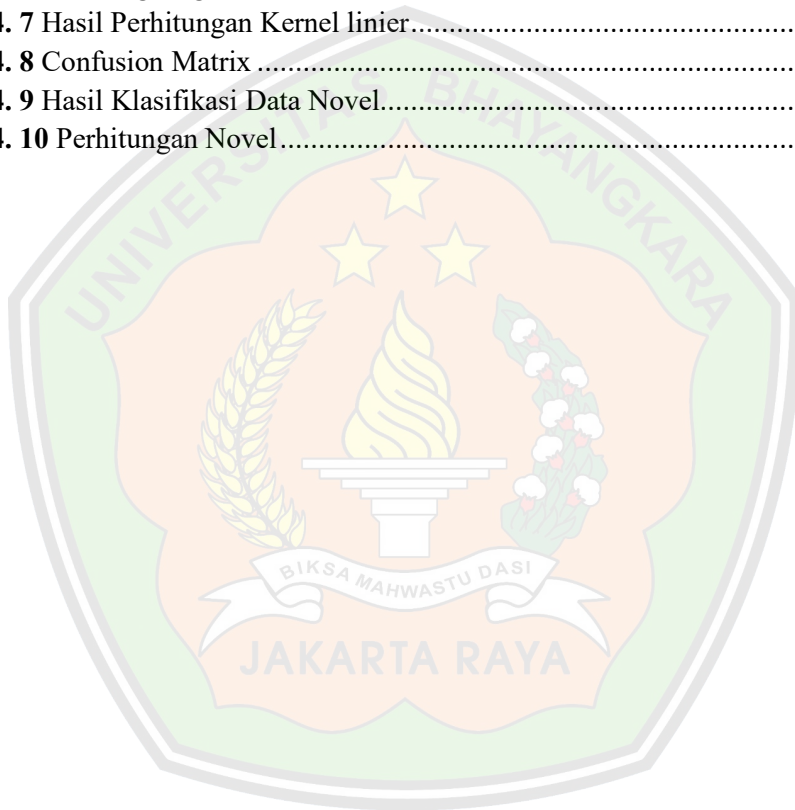
DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iii
ABSTRAK.....	iv
ABSTRACT	v
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vi
KARYA ILMIAH UNUK KEPENTINGAN AKADEMIK	vi
KATA PENGANTAR.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	5
1.3 Perumusan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
1.6 Batasan Penelitian	7
1.7 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA.....	8
2.1 Landasan Teori	8
2.2 Data Mining	12
2.2.1 Jenis-Jenis Data Mining	12
2.2.2 Tahapan CRISP-DM	14
2.3 Natural Language Processing (NLP).....	16
2.3.1 Tahap-Tahap Natural Language Processing (NLP).....	17
2.4 Support Vector Machine (SVM).....	18
2.4.1 Tahapan Support Vector Machine	19
2.5 Teori Slovin.....	23

2.5.1 Penerapan Slovin	24
2.6 Teori Novel.....	25
2.6.1 Jenis Jenis Genre Novel	25
2.7 Wattpad	27
BAB III_METODOLOGI PENELITIAN.....	30
3.1 Tipe Penelitian.....	30
3.1.1 Strategi dan Rancangan Penelitian.....	31
3.1.2 Kerangka Penelitian	31
3.1.3 Subjek atau Partisipan penelitian	33
3.2 Metode pengumpulan data.....	34
3.3 Metode analisis data	34
3.4 Model Data Mining	36
3.5 Peralatan Pendukung	37
BAB IV_HASIL DAN PEMBAHASAN.....	39
4.1 Hasil Penelitian	39
4.2 Tahapan Business Understanding.....	39
4.3 Tahapan Data Understanding	40
4.4 Tahapan Data Preparation	41
4.5 Tahapan Modeling	46
4.6 Tahapan Evaluasi.....	49
4.7 Tahapan Deployment.....	52
BAB V_PENUTUP.....	55
5.1 Kesimpulan	55
5.2 Keterbatasan.....	55
5.3 Saran	56
DAFTAR PUSTAKA	57
LAMPIRAN	60

DAFTAR TABEL

Tabel 1. 1 Pencapaian Wattpad.....	3
Tabel 2. 1 Penelitian Terdahulu.....	9
Tabel 4. 1 Data Novel	40
Tabel 4. 2 Tokenisasi	41
Tabel 4. 3 Normalisasi	42
Tabel 4. 4 Stemming	43
Tabel 4. 5 Stop word removal	44
Tabel 4. 6 Handling Angka	44
Tabel 4. 7 Hasil Perhitungan Kernel linier.....	48
Tabel 4. 8 Confusion Matrix	50
Tabel 4. 9 Hasil Klasifikasi Data Novel.....	52
Tabel 4. 10 Perhitungan Novel.....	53



DAFTAR GAMBAR

Gambar 1. 1 Pencapaian Wattpad [2]	2
Gambar 1. 2 Data Pengunjung Bulain Mei 2024 [3].....	3
Gambar 1. 3 Grafik Pengunjung Bulan Desember 2023 – Mei 2024 [3].....	4
Gambar 2. 1 Klasifikasi Mencari Hyperplane [22]	19
Gambar 2. 2 Jumlah Novel Genre Random [25].....	24
Gambar 3. 1 Kerangka Penelitian	32
Gambar 4. 1 TF-IDF.....	45
Gambar 4. 2 Labeling	46
Gambar 4. 3 Pembagian Data	46
Gambar 4. 4 Model SVM	47
Gambar 4. 5 Perhitungan Karna Linier	47
Gambar 4. 6 Evaluasi Klasifikasi.....	49



DAFTAR LAMPIRAN

Lampiran 1 Model Klasifikasi.....	61
Lampiran 2 Hasil Model.....	64
Lampiran 3 Plagiarisme	69
Lampiran 4 Biodata Mahasiswa.....	70
Lampiran 5 Kartu Bimbingan Skripsi.....	71

