

**PENERAPAN ALGORITMA WINNOWERING UNTUK
MENDETEKSI KEMIRIPAN JUDUL SKRIPSI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

DIMAS BAGUS PRASETYO

19.11.3020

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PENERAPAN ALGORITMA WINNOWER UNTUK
MENDETEKSI KEMIRIPAN JUDUL SKRIPSI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

DIMAS BAGUS PRASETYO

19.11.3020

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA WINNOWER UNTUK MENDETEKSI
KEMIRIPAN JUDUL SKRIPSI

yang disusun dan diajukan oleh

Dimas Bagus Prasetyo

19.11.3020

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Desember 2023

Dosen Pembimbing,



Dina Maulina, M.Kom

NIK. 190302250

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA WINNOWERING UNTUK MENDETEKSI
KEMIRIPAN JUDUL SKRIPSI

yang disusun dan diajukan oleh

Dimas Bagus Prasetyo

19.11.3020

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Desember 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bety Wulan Sari, M.Kom.
NIK. 190302254

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392

Dina Maulina, M.Kom.
NIK. 190302250

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Desember 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dimas Bagus Prasetyo
NIM : 19.11.3020

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma Winnowing untuk Mendeteksi Kemiripan Judul Skripsi

Dosen Pembimbing: Dina Maulina, M. Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Desember 2023

Yang Menyatakan,



5000
METERAI
TEMBEL
BCC42A9X739796188

Dimas Bagus Prasetyo

HALAMAN PERSEMBAHAN

Segala puji syukur, Alhamdulillah kehadiran Allah SWT yang Maha Kuasa, yang memberikan kemudahan dan kelancaran dalam setiap langkah yang saya ambil, salah satunya dapat menyelesaikan skripsi tepat pada waktunya. Oleh karena itu, semua hasil skripsi ini saya persembahkan skripsi ini untuk:

1. Terima kasih kepada kedua orang tua saya bapak Sudarmanto dan Ibu Suyatun yang telah membesarkan, membimbing saya dengan penuh kasih sayang, memberikan semangat dan selalu menjaga saya dalam naungan do'a di setiap langkah perjalanan saya.
2. Terima kasih kepada dosen pembimbing Ibu Dina Maulina, M.kom selaku dosen pembimbing saya, yang telah meluangkan waktunya untuk memberikan arahan, dan membimbing sehingga skripsi ini dapat terselesaikan dengan baik.
3. Terimakasih kepada Lutfiana Tazkiatun Nufus, perempuan yang selalu menemani proses saya selama ini, selalu memberikan semangat kepada saya sampai selesainya skripsi ini, dan membuat saya benar-benar bersyukur, benar-benar merasa bahagia memilikinya. Bismillah dengan kesungguhan hati kujadikan engkau pasangan sehidup surga.
4. Teman-teman yang selalu memberikan semangat dan support dalam proses penulisan skripsi ini. Serta semua pihak yang telah membantu dalam penelitian dan pengumpulan data untuk skripsi ini.

Terima kasih atas segala dukungan, doa, dan bantuan yang diberikan. Semoga hasil skripsi ini dapat bermanfaat bagi banyak orang dan menjadi amal jariah bagi kita semua. Amiin.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT, karena berkat limpahan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Algoritma Winnowing untuk Mendeteksi Kemiripan Judul Skripsi”. Skripsi ini ditulis untuk memenuhi salah satu persyaratan akademik perkuliahan demi memperoleh gelar Sarjana Komputer di Universitas AMIKOM Yogyakarta. Dalam penyusunan skripsi, penulis tak lepas dari pihak-pihak yang telah membantu dari awal hingga skripsi dapat terselesaikan dengan baik. Oleh sebab itu, penulis mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Dina Maulina, M.Kom. selaku Dosen Pembimbing yang telah memberikan arahan bagi penulis dalam menyelesaikan skripsi.
4. Kedua orangtua saya yang tiada henti memberikan doa dan dukungan sampai pada terselesainya skripsi ini.
5. Teman-teman yang selalu memberikan semangat dan support dalam proses penulisan skripsi ini.

Penulis menyadari bahwa di dalam skripsi ini masih terdapat kelemahan. Oleh sebab itu, penulis berharap adanya kritik dan saran demi perbaikan karya yang akan datang. Penulis mohon maaf apabila ada kesalahan kata yang kurang berkenan.

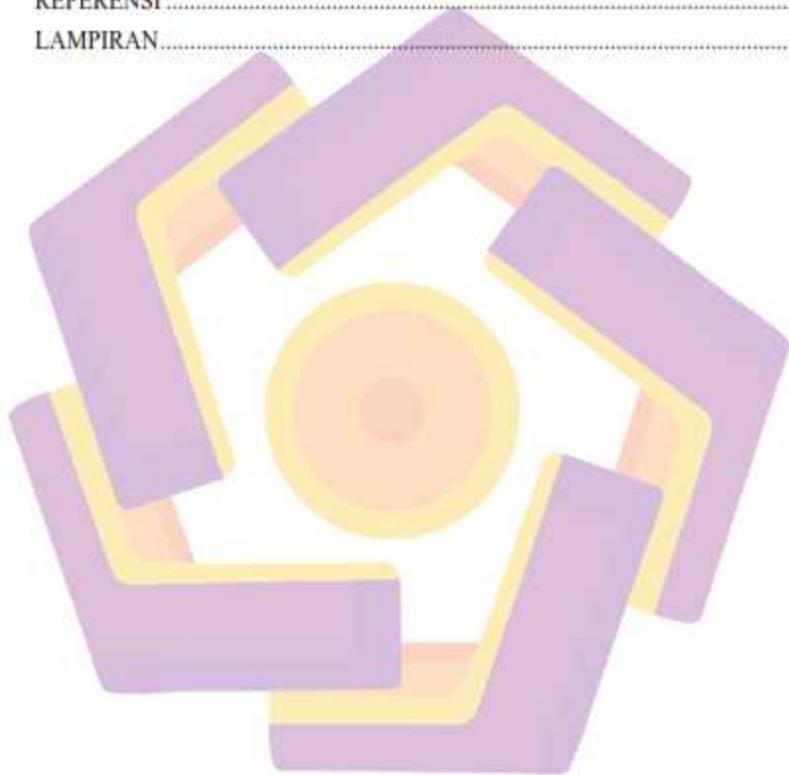
Yogyakarta, 20 Desember 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	10
BAB III METODE PENELITIAN	22
3.1 Objek Penelitian.....	22
3.2 Alat dan Bahan.....	22

3.3 Alur Penelitian	23
BAB IV HASIL DAN PEMBAHASAN	40
BAB V PENUTUP	53
5.1 Kesimpulan	53
5.2 Saran	53
REFERENSI	54
LAMPIRAN	57



DAFTAR TABEL

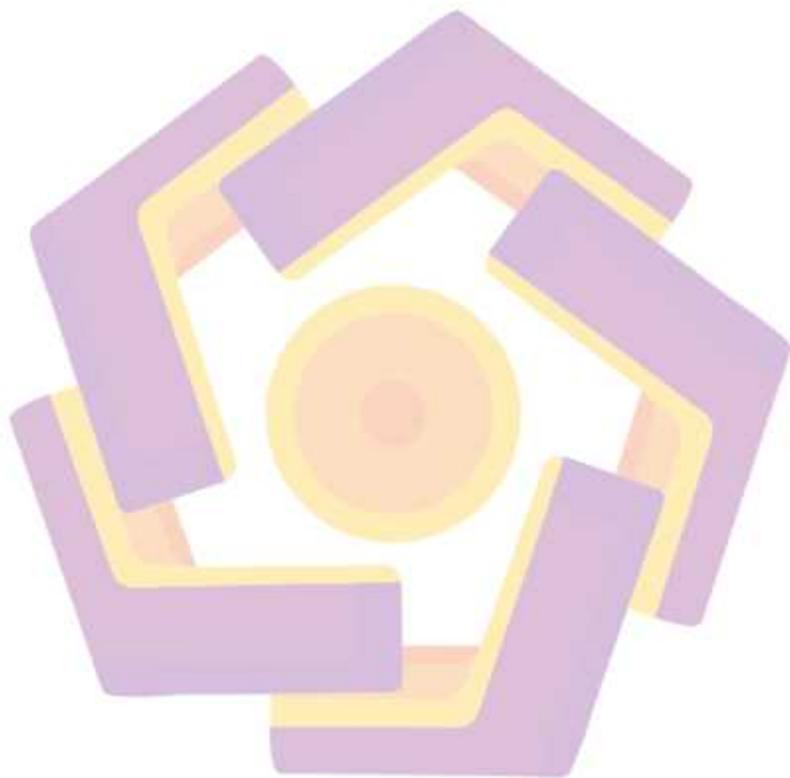
Tabel 2.1. Keaslian Penelitian	7
Tabel 2.2. <i>Flowchart</i> beserta artinya	11
Tabel 2.3. Pembentukan <i>N-gram</i>	19
Tabel 3.1. Spesifikasi Perangkat Keras	22
Tabel 3.2. Spesifikasi Perangkat Lunak	23
Tabel 3.3. Case Folding	27
Tabel 3.4. Remove Punctuation	27
Tabel 3.5. Remove Space	27
Tabel 3.6. Mengambil 2 Judul Untuk Dibandingkan	28
Tabel 3.7. Preprocessing Judul Pertama	29
Tabel 3.8. Preprocessing Judul Kedua	29
Tabel 3.9. Pembentukan <i>N-Gram</i> Judul 1	30
Tabel 3.10. Pembentukan <i>N-Gram</i> Judul 2	30
Tabel 3.11. Nilai <i>Hash</i> Judul 1 dan 2	31
Tabel 3.12. Pembentukan <i>Window</i> Judul 1 dan 2	32
Tabel 3.13. <i>Fingerprint</i> Judul 1 dan 2	35
Tabel 3.14. Tingkat Plagiarisme	37
Tabel 3.15. Pengujian dengan <i>N-Gram</i> dan <i>Window</i> sama	37
Tabel 3.16. Pengujian dengan <i>N-Gram</i> dan <i>Window</i> sama	38
Tabel 4.1. Hasil <i>Black Box Testing</i>	52

DAFTAR GAMBAR

Gambar 2.1. Gambar ASCII Tabel	18
Gambar 3.1. Alur Penelitian	24
Gambar 3.2. Perancangan ERD	25
Gambar 3.3. Mapping Table	25
Gambar 3.4. Alur Proses <i>Preprocessing Data</i>	26
Gambar 3.5. Grafik hasil uji coba dengan <i>n-gram</i> dan <i>window</i> sama	37
Gambar 3.6. Grafik pengujian dengan <i>n-gram</i> dan <i>window</i> berbeda	38
Gambar 4.1. Kode program lowercase	40
Gambar 4.2. Kode untuk menghapus spasi	41
Gambar 4.3. Kode untuk menghapus tanda baca	41
Gambar 4.4. Variabel Pembentukan <i>N-Gram</i>	42
Gambar 4.5. Kode fungsi pembuatan <i>N-Gram</i>	42
Gambar 4.6. Variabel untuk menampung nilai <i>hash</i>	42
Gambar 4.7. Kode program fungsi <i>rolling hash</i>	43
Gambar 4.8. Variabel untuk menampung <i>window</i>	43
Gambar 4.9. Kode program fungsi pembuatan <i>window</i>	43
Gambar 4.10. Variabel untuk menampung <i>fingerprint</i>	44
Gambar 4.11. Fungsi pembuatan <i>fingerprint</i>	44
Gambar 4.12. Variabel untuk menampung <i>intersection</i>	44
Gambar 4.13. Kode untuk menghitung <i>intersection</i>	45
Gambar 4.14. Variabel untuk menampung <i>similarity</i>	45
Gambar 4.15. Kode untuk menghitung persentase kemiripan	45
Gambar 4.16. Tampilan Halaman Login	46
Gambar 4.17. Tampilan Halaman Beranda	47
Gambar 4.18. Tampilan Halaman Data Judul	48
Gambar 4.19. Tampilan Halaman Cek Kemiripan	49
Gambar 4.20. Hasil Cek Kemiripan	50
Gambar 4.21. Detail Cek Kemiripan	51

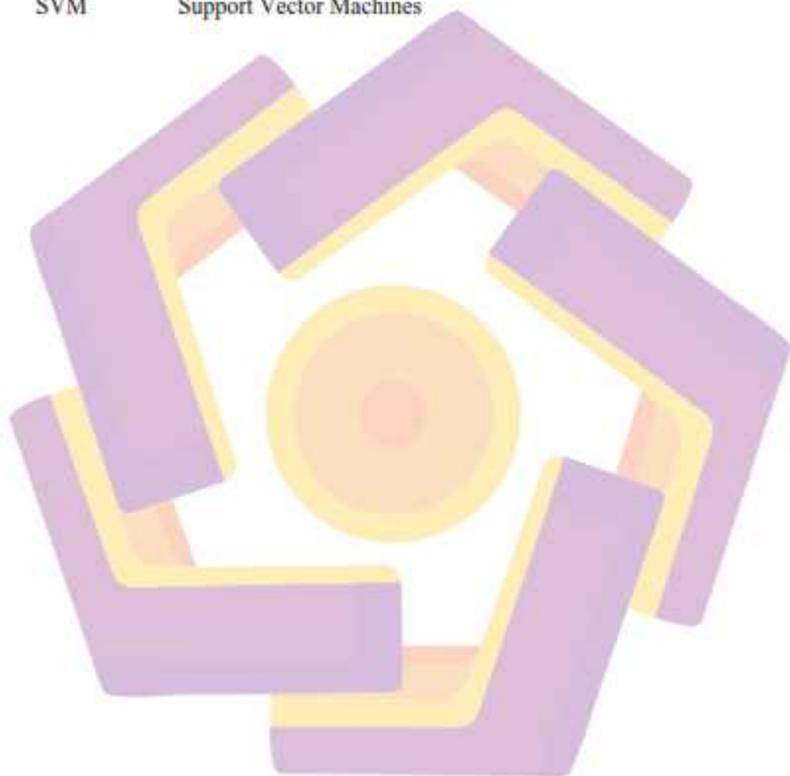
DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



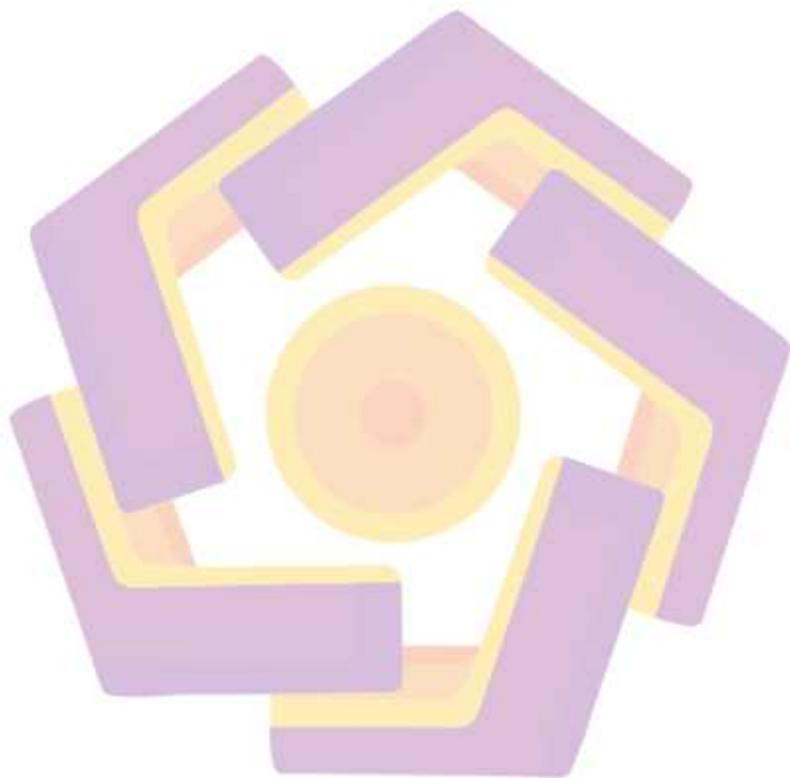
DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines



DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Eigen Value	akar akar persamaan



INTISARI

Dalam dunia pendidikan tinggi terdapat istilah skripsi yaitu tempat untuk menuangkan karya tulis ilmiah dari hasil penelitian S1 dengan pembahasan mengenai ilmu tertentu berdasarkan kajian pustaka oleh para ahli, hasil penelitian lapangan atau hasil pengembangan eksperimen. Setiap tahunnya Universitas AMIKOM Yogyakarta meluluskan mahasiswa yang membuat banyak skripsi yang telah dibuat. Seiring berjalannya waktu semakin banyak judul skripsi yang dibuat yang memungkinkan judul yang diajukan dan judul yang terdahulu memiliki kemiripan. Hal tersebut memungkinkan terjadi kesamaan dalam topik yang dibahas dalam skripsi. Untuk mengatasi hal tersebut dibutuhkan suatu metode untuk mendeteksi kemiripan judul skripsi dengan suatu algoritma yang dapat diterapkan.

Metode yang dapat digunakan untuk mendeteksi kemiripan teks adalah algoritma *winnowing*. Algoritma *winnowing* digunakan sebagai pencarian dokumen *fingerprinting* yaitu metode untuk mendeteksi keakuratan salinan dari kalimat atau dokumen. Secara garis besar penerapan algoritma *winnowing* untuk mendeteksi kemiripan teks adalah membentuk *n-gram*, pembentukan *hashing*, pembentukan *window*, pencarian *fingerprinting*, dan mencari persentase dengan *jaccard coefficient*.

Hasil akhir dari penerapan algoritma *winnowing* ini adalah suatu *prototype* sistem sederhana yang memiliki fitur untuk mendeteksi kemiripan judul skripsi yang dapat digunakan oleh dosen pembimbing untuk dapat mengecek judul yang telah diajukan mahasiswa. Jika judul yang diajukan banyak kemiripan dengan judul sebelumnya maka dosen pembimbing dapat merekomendasikan judul lain namun jika judul yang diajukan tidak banyak kemiripan maka dapat disetujui lalu dapat dimasukkan ke database judul.

Kata kunci: Algoritma *Winnowing*, Skripsi, *n-gram*, *Fingerprinting*, *hashing*

ABSTRACT

In the world of higher education there is the term thesis, which is a place to pour scientific papers from undergraduate research with a discussion of certain knowledge based on literature review by experts, field research results or experimental development results. Every year AMIKOM Yogyakarta University graduates students who make many theses that have been made. Over time, more and more thesis titles are made which allows the proposed title and the previous title to have similarities. This allows similarities in the topics discussed in the thesis. To overcome this, a method is needed to detect the similarity of thesis titles with an algorithm that can be applied.

The method that can be used to detect text similarity is the winnowing algorithm. The winnowing algorithm is used as a fingerprinting document search, which is a method to detect the accuracy of copies of sentences or documents. Broadly speaking, the application of the winnowing algorithm to detect text similarity is forming n-grams, hashing formation, window formation, fingerprinting search, and finding the percentage with jaccard coefficient.

The final result of the application of this winnowing algorithm is a simple prototype system that has a feature to detect similarities in thesis titles that can be used by supervisors to be able to check the titles that have been submitted by students. If the proposed title has many similarities with the previous title, the supervisor can recommend another title, but if the proposed title does not have many similarities, it can be approved and then can be entered into the title database.

Keyword: *Winnowing Algorithm, Thesis, N-gram, Fingerprinting, Hashing*