

Analisa Perbandingan Algoritma *Levenshtein Distance* dan *Cosine Similarity* pada Kasus Plagiarisme

*Diajukan Untuk Menyusun Tugas Akhir
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

SARNIKO MARWANSAH

NIM : 09121002049

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2019

LEMBAR PENGESAHAN TUGAS AKHIR

**Analisa Perbandingan Algoritma *Levenshtein Distance*
dan *Cosine Similarity* pada Kasus Plagiarisme**

Oleh :

SARNIKO MARWANSAH
NIM : 09121002049

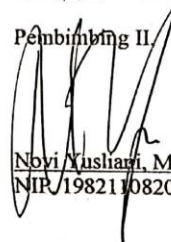
Indrajaya, Juli 2019

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002


Pembimbing II,



Novi Yushanti, M. T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,




Rifkie Primartha, M. T.
NIP. 197704012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 31 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Sarniko Marwansah

NIM : 09121002049

Judul : Analisa Perbandingan Algoritma *Levenshtein Distance* dan *Cosine Similarity* pada Kasus Plagiarisme

1. Pembimbing I

M. Fachrurrozi, M.T.

NIP. 198005222008121002

2. Pembimbing II

Novi Yusliani, M.T.

NIP. 198211082012122001

3. Penguji I

Ali Ibrahim, M. T.

NIP. 198407212010031201

4. Penguji II

M. Naufal Rachmatullah, M.T.

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M. T.

NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Sarniko Marwansah

NIM : 09121002049

Program Studi : Teknik Informatika

Judul Skripsi : Analisa Perbandingan Algoritma *Levenshtein Distance* dan *Cosine Similarity* pada Kasus Plagiarisme

Hasil Pengecekan Software *iThenticate Turnitin* : 10 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Juli 2019




Sarniko Marwansah
NIM. 09121002049

Motto :

- *Perhaps you hate a thing and it is good for you, and perhaps you love a thing and it is bad for you. And Allah Knows, while you know not.*
- *sesungguhnya beserta kesulitan itu ada kemudahan.*

Kupersembahkan karya tulis ini kepada :

- *Kedua orangtuaku*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer Universitas Sriwijaya*

COMPARATIVE ANALYSIS OF *LEVENSHEIN DISTANCE* AND *COSINE SIMILARITY* ALGORITHM IN PLAGIARISM CASE

By:

Sarniko Marwansah

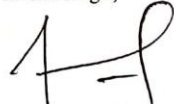
09121002049

ABSTRACT

Plagiarism is a crime committed by stealing or taking someone's work or ideas which are then made as if his own. In overcoming it requires an effective method, namely the automatic method. In this research the *Levenshtein Distance* and *Cosine Similarity* methods are used to detect plagiarism. The results of the comparison of the two methods found that the two methods namely the *Levenshtein Distance* and *Cosine Similarity* methods have the same level of accuracy based on 50 times the test. The results of the average percentage of similarity from the detection of text similarity using the *Levenshtein Distance* and *Cosine Similarity* algorithms in this study were 100% at 100% *actual similarities*, 19.81% at 80% *actual similarities*, 39.86% at 60% *actual similarities*, 60.04% in *actual similarities* 40%, and 79.85% in *actual similarities* 20%.

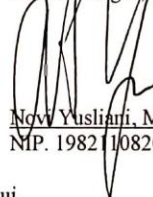
Keywords: Plagiarism, *Levenshtein Distance*, *Cosine Similarity*.

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002

Indralaya, Juli 2019
Pembimbing II,

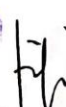


Ngvy Yusliani, M. T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika,




Rifkie Primurtha, M. T.
NIP. 197706012009121004

ANALISA PERBANDINGAN ALGORITMA *LEVENSHTEIN DISTANCE* DAN *COSINE SIMILARITY* PADA KASUS PLAGIARISME

Oleh :

Sarniko Marwansah

09121002049

ABSTRAK

Plagiarisme merupakan tindak kejahatan yang dilakukan dengan mencuri atau mengambil karya atau ide seseorang yang kemudian dijadikan seolah-olah miliknya. Dalam penanggulangannya dibutuhkan metode yang efektif, yaitu metode otomatis. Dalam penelitian ini digunakan metode *Levenshtein Distance* dan *Cosine Similarity* untuk mendeteksi plagiarisme. Hasil dari perbandingan kedua metode didapatkan bahwa kedua metode yaitu metode *Levenshtein Distance* dan *Cosine Similarity* memiliki tingkat akurasi yang sama berdasarkan 50 kali pengujian. Hasil rata-rata persentase kemiripan dari pendeteksian kemiripan teks menggunakan algoritma *Levenshtein Distance* dan *Cosine Similarity* dalam penelitian ini sebesar 100% pada *actual similarities* 100%, 19,81% pada *actual similarities* 80%, 39,86% pada *actual similarities* 60%, 60,04% pada *actual similarities* 40%, dan 79,85% pada *actual similarities* 20%.

Kata Kunci : Plagiarisme, *Levenshtein Distance*, *Cosine Similarity*.

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002

Indralaya, Juli 2019

Pembimbing II,



Novi Yuslim, M. T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M. T.
NIP. 197706012009121004

KATA PENGANTAR

Bismillahirrahmanir rahiim

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Nurzali dan Armini, serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M. T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika dan Ibu Hardini Novianti, M. T. selaku Sekretaris Jurusan Teknik Informatika.
3. Bapak M. Fachrurrozi, M.T selaku dosen pembimbing I dan Ibu Novi Yusliani, M.T selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Ibu Desty Rodiah, M.T. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Ali Ibrahim, M. T. selaku dosen penguji I, dan Bapak Naufal Rahmatullah selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Abson, Aud, Genda, Putri, Dara, serta seluruh teman-teman jurusan Teknik Informatika yang telah saling berbagi selama masa perkuliahan ini.
9. BPH HMIF, Wifi, Digital Creative yang telah memberikan ruang bagi Penulis untuk berprestasi dan berkarya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Juli 2019

Sarniko Marwansah

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II TINJAUAN PUSTAKA	
2.1 Pendahuluan....	II-1
2.2 Plagiarisme.....	II-1

2.3	<i>Preprocessing</i>	II-3
2.4.	<i>Levenshtein Distance</i>	II-5
2.5	<i>Cosine Similarity</i>	II-8
2.6	<i>Term Frequency-Inverse Document Frequency</i>	II-9
2.7	<i>Rational Unified Process(RUP)</i>	II-11
	2.7.1 <i>Fase-Fase Rational Unified Process (RUP)</i>	II-12
2.8	Penelitian Lain Yang Relevan.....	II-13

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
	3.2.1 Jenis Data	III-1
	3.2.2 Sumber Data	III-1
	3.2.3 Metode Pengumpulan Data.....	III-2
3.3	Tahapan Penelitian.....	III-3
	3.3.1 Menetapkan Kerangka Kerja/Framework	III-3
	<i>a. Preprocessing</i>	III-4
	<i>b. Levenshtein Distance</i>	III-5
	<i>c. Pembobotan TF-IDF</i>	III-6
	<i>d. Perhitungan Kemiripan teks menggunakan Cosine Similarity</i>	III-7
	3.3.2 Menetapkan Kriteria Pengujian	III-8
	3.3.3 Menentukan Format Data Pengujian	III-8
	3.3.4 Menentukan Alat yang digunakan dalam Pelaksanaan Penelitian	III-9
	3.3.5 Melakukan Pengujian Penelitian.....	III-10
3.4	Metode Pengembangan Perangkat Lunak	III-10
3.5	Manajemen Proyek Penelitian	III-14

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan.....	IV-1
4.2	Fase Insepsi.....	IV-1
4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.2.2.1	Fitur Perhitungan Persentase Kemiripan Teks	IV-2
4.2.3	Analisis Dan Desain	IV-4
4.2.3.1	Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.2	Analisis Data.....	IV-5
4.2.3.3	Analisis Preprocessing.....	IV-7
4.2.3.3.1	Analisis Case Folding.....	IV-7
4.2.3.3.2	Analisis Tokenizing.....	IV-8
4.2.3.3.3	Analisis Stopword Removal	IV-9
4.2.3.3.4	Analisis Stemming.....	IV-10
4.2.3.4	Analisa Perbandingan Algoritma Levenshtein Distance dan Cosine Similarity	IV-11
4.2.3.5	Desain Perangkat Lunak.....	IV-12
4.2.3.6	Model Use Case.....	IV-12
4.3	Fase Elaborasi	IV-20
4.3.1	Pemodelan Bisnis	IV-20
4.3.1.1	Perancangan Data	IV-20
4.3.1.2	Perancangan Antarmuka.....	IV-21
4.3.2	Kebutuhan Sistem	IV-22
4.3.3	Sequence Diagram.....	IV-23
4.4	Fase Konstruksi.....	IV-24
4.4.1	Diagram Kelas.....	IV-25
4.4.2	Implementasi	IV-25
4.4.2.1	Implementasi Kelas	IV-26
4.4.2.2	Implementasi Antarmuka	IV-27

4.5	Fase Transisi	IV-27
4.5.1	Pemodelan Bisnis	IV-28
4.5.2	Kebutuhan Sistem	IV-28
4.5.3	Rencana Pengujian	IV-28
4.5.4	Implementasi	IV-30
BAB V HASIL DAN ANALISIS PENELITIAN		
5.1	Pendahuluan.....	V-1
5.2	Data Hasil Pengujian	V-1
5.2.1	Konfigurasi Pengujian I	V-2
5.2.2	Konfigurasi Pengujian II	V-3
5.2.3	Hasil Pengujian	V-4
5.3	Analisis Hasil Pengujian.....	V-5
BAB VI KESIMPULAN DAN SARAN		
6.1	Kesimpulan	VI-1
6.2	Saran	VI-1
DAFTAR PUSTAKA		xviii
LAMPIRAN.....		xx

DAFTAR TABEL

	Halaman
Tabel II-1 Hasil Perhitungan <i>Cosine Similarity</i>	II-8
Tabel II-2 Hasil Perhitungan <i>Tf-idf</i>	II-10
Tabel III-1 Inisialisasi <i>Levenshtein</i>	III-1
Tabel III-2 Matriks Nilai dari Dua Dokumen Teks	III-6
Tabel III-3 Rancangan Hasil Pendeteksian Kemiripan Teks	III-8
Tabel III-4 Pengembangan Perangkat Lunak dengan Metode RUP	III-11
Tabel III-5 Penjadwalan Penelitian dalam bentuk <i>Word Breakdown Structure</i> (WBS)	III-15
Tabel IV-1 Kebutuhan Fungsional Perangkat Lunak.....	IV-4
Tabel IV-2 Kebutuhan Non-Fungsional Perangkat Lunak	IV-4
Tabel IV-3 Teks Dokumen Sumber dan Dokumen Pembanding.....	IV-8
Tabel IV-4 Teks Dokumen Sumber dan Dokumen Pembanding (Hasil Case Folding)	IV-8
Tabel IV-5 Teks Dokumen Sumber dan Dokumen Pembanding (Sebelum Tokenizing)	IV-9
Tabel IV-6 Teks Dokumen Sumber dan Dokumen Pembanding (Setelah Tokenizing)	IV-9
Tabel IV-7 Teks Dokumen Sumber dan Dokumen Pembanding (Sebelum Stopword Removal)	IV-10
Tabel IV-8 Teks Dokumen Sumber dan Dokumen Pembanding (Setelah Stopword Removal)	IV-10
Tabel IV-9 Teks Dokumen Sumber dan Dokumen Pembanding (Sebelum Steaming)	IV-11
Tabel IV-10 Teks Dokumen Sumber dan Dokumen Pembanding (Setelah Steaming)	IV-11

Tabel IV-11 Definisi Aktor.....	IV-13
Tabel IV-12 Deskripsi Use Case.....	IV-14
Tabel IV-13 Skenario Use Case Memasukan Dokumen Uji	IV-15
Tabel IV-14 Skenario Use Case Memasukan Dokumen Sumber	IV-16
Tabel IV-15 Skenario Use Case Melakukan Pendeteksian Kemiripan Teks Menggunakan Algoritma <i>Levenshtein Distance</i> dan <i>Cosine Similarity</i> IV-17	IV-17
Tabel IV-16 Tabel Implementasi Kelas	IV-27
Tabel IV-17 Rencana Pengujian Use Case Menghitung Persentase Kemiripan Teks Menggunakan Algoritma <i>Levenshtein Distance</i> dan <i>Cosine Similarity</i> IV-30	IV-30
Tabel V-1 Informasi Dokumen Sumber.....	V-1
Tabel V2 Informasi Dokumen Uji (Datasets).....	V-2
Tabel V-3 Hasil Pengujian Pendeteksian Kemiripan Teks dengan Algoritma <i>Levenshtein Distance</i>	V-3
Tabel V-4 Hasil Pengujian Pendeteksian Kemiripan Teks dengan Algoritma <i>Cosine Similarity</i>	V-4
Tabel V-5 Rata-Rata Persentase Kemiripan Setiap Algoritma	V-5

DAFTAR GAMBAR

	Halaman
Gambar II-1 Arsitektur <i>Rational Unified Process</i>	II-12
.Gambar III-1 Diagram Tahapan Perangkat Lunak.....	III-4
Gambar IV-1 Diagram Proses Kerja Pendeteksian Kemiripan Teks	IV-3
Gambar IV-2 Contoh Data Kamus Stopword	IV-7
Gambar IV-3 Diagram Use Case Perangkat Lunak Perhitungan Kemiripan Teks.....	IV-13
Gambar IV-4 Kelas Analisis Browse Dokumen Uji.....	IV-18
Gambar IV-5 Kelas Analisis Dokumen Sumber.....	IV-19
Gambar IV-6 Kelas Analisis Levenshtein.....	IV-19
Gambar IV-7 Kelas Analisis Cosine	IV-20
Gambar IV-8 Rancangan Antarmuka Perangkat Lunak	IV-22
Gambar IV-9 Sequence Diagram Dokumen Uji	IV-24
Gambar IV-10 Sequence Diagram Dokumen Sumber	IV-24
Gambar IV-11 Sequence Diagram Calculate.....	IV-25
Gambar IV-12 Kelas Diagram	IV-26
Gambar IV-13 Antarmuka Perangkat Lunak.....	IV-28

DAFTAR LAMPIRAN

	Halaman
LAMPIRAN 1 Source Code Program	L-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2 Latar Belakang

Pengaruh dari perkembangan teknologi telah memberikan banyak kemudahan kepada berbagai bidang termasuk juga dalam pencarian informasi. Karena kemudahan-kemudahan dalam mencari informasi sehingga menimbulkan sisi buruk dari sebagian orang dalam bidang akademik khususnya dalam penulisan karya ilmiah seperti plagiarisme. Plagiarisme adalah pengambilan karangan atau pendapat orang lain secara keseluruhan atau sebagian tanpa mencantumkan sumber referensi dan menjadikannya seolah-olah karangan atau pendapat sendiri.

Agar dapat mengetahui keaslian karya tulis seseorang dibutuhkan pemeriksaan terhadap karya itu sendiri. Salah satu caranya adalah dengan metode manual yaitu dengan memeriksa karya itu dengan seksama. Namun hal ini tidaklah efektif dan memakan banyak waktu jika teks yang akan dibandingkan sangatlah banyak. Sehingga dibutuhkan suatu program yang dapat melakukan pendeteksian keaslian karya tulis secara otomatis.

Beberapa penelitian tentang plagiarisme telah banyak dilakukan. Salah satu algoritma yang digunakan adalah algoritma *Levenshtein Distance*. Pada penelitian yang dilakukan oleh Ajie & Bangsa (2014) menggunakan algoritma *Levenshtein Distance* dalam menghitung persentase kemiripan teks, dilakukan pengujian sebanyak dua kali dengan data uji yang berbeda. Data uji 1 dilakukan dengan membandingkan dua dokumen yang sama tetapi telah diubah letak struktur kalimatnya seperti mengubah paragraf induktif menjadi deduktif atau sebaliknya. Data uji 2 dilakukan dengan membandingkan dua dokumen yang struktur kata didalam kalimatnya telah diubah, yaitu dengan mengubah kalimat pasif menjadi aktif atau sebaliknya. Hasil persentase kemiripan teks yang didapat dari data uji 1 adalah 80.87% sedangkan dari data uji 2 adalah 58.81%. Hasil pengujian menunjukkan bahwa, dengan tahap *preprocessing* yang digunakan pada penelitian ini, program yang mengimplementasi algoritma *Levenshtein Distance* lebih cocok digunakan untuk mendeteksi kemiripan teks dengan cara kata demi kata (*word for word plagiarizing*).

Selain itu ada juga algoritma *Cosine Similarity*. Pada penelitian yang dilakukan oleh Wahyuni, Prastiyanto dan Suprpto (2017) menggunakan algoritma *Cosine Similarity* dan pembobotan *tf-idf* untuk mengklasifikasi dokumen skripsi. Dilakukan 2 tahap pengujian, yaitu tahap *training* data dan tahap *testing*. Pada tahap *training*, data yang digunakan telah diketahui kategorinya. Pada tahap *testing*, dimana data-data yang diolah belum diketahui kategorinya akan mendapatkan kategori secara otomatis oleh sistem. Hasil

pengujian didapatkan persentase rata-rata hasil pengujian oleh sistem sebesar 98%.

Berdasarkan keterangan diatas, dapat disimpulkan bahwa algoritma *Levenshtein Distance* dan *Cosine Similarity* dapat digunakan untuk melakukan pendeteksian kemiripan teks. Penelitian yang akan dilakukan yaitu membandingkan persentase kemiripan untuk satu kali perhitungan kemiripan teks dan kecepatan proses antara algoritma *Levenshtein Distance* dan *Cosine Similarity* yang dijalankan oleh sistem. Untuk mengetahui algoritma mana yang memiliki tingkat persentase kemiripan yang lebih besar.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan , rumusan masalah pada penelitian ini adalah bagaimana perbandingan tingkat persentase algoritma *Levenshtein Distance* dan *Cosine Similarity* pada sebuah sistem pendeteksian plagiarisme.

Untuk menyelesaikan permasalahan yang disebutkan diatas, maka dalam penelitian ini dibagi dalam beberapa *research question* sebagai berikut:

1. Bagaimana mekanisme *Levenshtein Distance* dalam sistem pendeteksian plagiarisme?
2. Bagaimana mekanisme *Cosine Similarity* dalam sistem pendeteksian plagiarisme?

3. Bagaimana hasil perbandingan *Levenshtein Distance* dan *Cosine Similarity* pada sistem pendeteksian plagiarisme berdasarkan dari hasil persentase?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma *Levenshtein Distance* dan *Cosine Similarity* ke dalam sistem pendeteksian plagiarisme.
2. Mengetahui hasil persentase kemiripan dari algoritma *Levenshtein Distance* dan *Cosine Similarity* yang diterapkan kedalam sistem.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah sebagai berikut :

1. Memahami Algoritma *Levenshtein Distance* dan *Cosine Similarity* sebagai metode pada pendeteksian plagiarisme.
2. Hasil penelitian dapat dijadikan sebagai rujukan bagi peneliti lain dalam mengembangkan sistem pendeteksian plagiarisme.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data set yang akan diproses adalah dokumen berbahasa Indonesia.
2. Dokumen teks berekstensi .txt.
3. Sistem tidak memperhatikan kesalahan ejaan/penulisan pada dokumen

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Bab I ini menguraikan latar belakang, perumusan masalah, tujuan dan manfaat penelitian serta batasan masalah, dan sistematika penulisan penelitian.

BAB II. KAJIAN LITERATUR

Pada bab II ini akan membahas landasan teori yang digunakan dalam penelitian ini, seperti analisis *preprocessing*, Algoritma *Levenshtein Distance*, Algoritma *Cosine Similarity*, dan metode pengembangan perangkat lunak yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab III ini akan membahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Pada tahapan penelitian ini akan dibahas dengan rinci dengan mengacu pada suatu kerangka kerja. Serta pada bagian akhir bab ini akan dijabarkan perancangan manajemen proyek perangkat lunak untuk pelaksanaan penelitian ini.

1.8 Kesimpulan

Pada penelitian ini berfokus pada deteksi plagiarisme ekstrinsik. Metode yang digunakan untuk mendeteksi plagiarisme ini adalah *Levenshtein Distance* dan *Cosine Similarity*. Pengujian akan dilakukan dengan melihat persentase keakurasian hasil kemiripan teks dari sistem

pendeteksian plagiarisme dengan metode *Cosine Similarity*. Kemudian hasil pengujian akan dibandingkan dengan hasil dari menggunakan metode *Levenshtein Distance*.

DAFTAR PUSTAKA

- Boisvert, R. F. and Irwin, M. J. (2006) 'Plagiarism on the rise', *Communications of the ACM*, 49(6), p. 23. doi: 10.1145/1132469.1132487.
- Baruah, D., & Kakoti Mahanta, A. (2013). A New Similarity Measure with Length Factor for Plagiarism Detection. *International Journal of Computer Applications*, 72(14), 14–17. <https://doi.org/10.5120/12561-8671>
- Clough, P. (2003). Old and new challenges in automatic plagiarism detection. *National Plagiarism Advisory Service*.
- Gusmita, R. H., Durachman, Y., Harun, S., Firmansyah, A. F., Sukmana, H. T., & Suhaimi, A. (2014). A rule-based question answering system on relevant documents of Indonesian Quran Translation. *2014 International Conference on Cyber and IT Service Management, CITSM 2014*. <https://doi.org/10.1109/CITSM.2014.7042185>.
- Firdaus, A., Ernawati and Vatesia, A. (2014) 'Aplikasi Pendeteksi Kemiripan pada Dokumen Teks Menggunakan Algoritma Nazief & Andriani Dan Metode Cosine Similarity', *Jurnal Teknologi Informasi*, 10(April), pp. 96–109.
- Krisnawati, L. D., & Schulz, K. U. (2014). *Plagiarism Detection for Indonesian Texts*. <https://doi.org/10.1145/2539150.2539213>
- Imbar, R. V. et al. (no date) 'Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks', pp. 31–42.
- Ryansyah, A. and Andayani, S. (2017) 'Implementasi Algoritma TF-IDF Pada

- Pengukuran Kesamaan Dokumen', *Jurnal Sistem & Teknologi Informasi Komunikasi*, 1(1), pp. 1–10.
- Joy, M., & Luck, M. (1999). Plagiarism in programming assignments. *IEEE Transactions on Education*. <https://doi.org/10.1109/13.762946>
- Martin, B. (1994). Plagiarism: A Misplaced Emphasis, by Brian Martin. *Information Ethics*. Retrieved from <https://www.uow.edu.au/~bmartin/pubs/94jie.html>
- Osman, A. H., Salim, N., Binwahlan, M. S., Hentably, H., & Ali, A. M. (2011). Conceptual similarity and graph-based method for plagiarism detection. *Journal of Theoretical and Applied Information Technology*.
- Pressman, R. S., & Maxim, B. R. (2015). Software Engineering: A Practitioner's Approach, Eighth Edition. In *ACM SIGSOFT Software Engineering Notes*. <https://doi.org/10.1145/1226816.1226822>
- Stein, B., & zu Eissen, S. M. (2006). *Near Similarity Search and Plagiarism Analysis*. https://doi.org/10.1007/3-540-31314-1_52
- Su, Z., Ahn, B. R., Eom, K. Y., Kang, M. K., Kim, J. P., & Kim, M. K. (2008). Plagiarism detection using the Levenshtein distance and Smith-Waterman algorithm. *3rd International Conference on Innovative Computing Information and Control, ICICIC'08*. <https://doi.org/10.1109/ICICIC.2008.422>