

**EVALUASI PENGUKURAN JARAK PADA ALGORITMA WINNOWER
UNTUK MENDETEKSI PLAGIASI BILINGUAL**

SKRIPSI

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Strata Satu (S1)
Pada Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Ponorogo



WAHIDUN NI'AM FITHRIONO ASYROFI
(16532569)

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH PONOROGO**

2019

HALAMAN PENGESAHAN

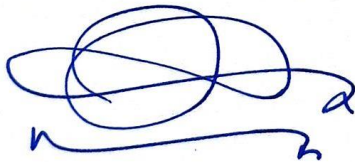
Nama : Wahidun Ni'am Fithriono Asyrofi
NIM : 16532569
Program Studi : Teknik Informatika
Fakultas : Teknik
Judul Skripsi : "Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual"

Isi dan formatnya telah disetujui dan dinyatakan memenuhi syarat untuk melengkapi persyaratan guna memperoleh Gelar Sarjana pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Ponorogo.

Ponorogo, Agustus 2020

Menyetujui,

Dosen Pembimbing I,



Dra. Ida Widaningrum, M.Kom

NIK. 19660417 201101 13

Dosen Pembimbing II,



Dyah Mustikasari, S.T., M.Eng

NIK. 19871007 201609 13

Mengetahui,

Dekan Fakultas Teknik,



Dr. Ir. Aliyadi, M.M., M.Kom

NIK. 19640103 19909 12

Ketua Program Studi Teknik Informatika,



Dyah Mustikasari, S.T., M.Eng

NIK. 19871007 201609 13

PERNYATAAN ORISINALITAS SKRIPSI

Yang bertandatangan dibawah ini :

Nama : Wahidun Ni'am Fithriono Asyrofi

NIM : 16532569

Program Studi : Teknik Informatika

Dengan ini menyatakan bahwa Skripsi saya dengan judul : "Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual" bahwa berdasarkan hasil penelusuran berbagai karya ilmiah, gagasan, dan masalah ilmiah yang saya rancang / teliti di dalam Naskah Skripsi ini adalah asli dari pemikiran saya. Tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam Naskah Skripsi ini dapat dibuktikan terdapat unsur- unsur plagiarisme, saya bersedia Ijazah saya dibatalkan, serta diproses sesuai dengan peraturan perundang- undangan yang berlaku.

Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar- benarnya.

Ponorogo, Agustus 2020



Wahidun Ni'am Fithriono Asyrofi

NIM. 16532569

HALAMAN BERITA ACARA UJIAN

Nama : Wahidun Ni'am Fithriono Asyrofi
NIM : 16532569
Program Studi : Teknik Informatika
Fakultas : Teknik
Judul Skripsi : "Evaluasi Pengukuran Jarak Pada Algoritma Winnowing
Untuk Mendeteksi Plagiasi Bilingual"

Telah diuji dan dipertahankan dihadapan Dosen penguji tugas akhir jenjang Strata
Satu (S1) pada :

Hari : Rabu
Tanggal : 5 Agustus 2020
Nilai : A

Dosen Penguji,

Dosen Penguji I,



Dr. Aslan Alwi, S.SI., M.Cs
NIK. 19720324 201101 13

Dosen Penguji II,



Indah Puji Astuti, S.Kom., M.Kom
NIK. 19860424 201609 13

Mengetahui,

Dekan Fakultas Teknik,



Dr. Ir. Aliyadi, M.M., M.Kom
NIK. 19640103 19909 12





Ketua Program Studi Teknik Informatika,



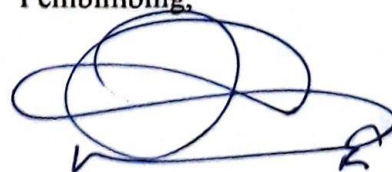
Dyah Mustikasari, S.T., M.Eng
NIK. 19871007 201609 13

**BERITA ACARA
BIMBINGAN SKRIPSI**

1. Nama : Wahidun Ni'am Fithriono Asyrofi
2. NIM : 16532569
3. Program Studi : Teknik Informatika
4. Fakultas : Teknik
5. Judul Skripsi : Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual
6. Dosen Pembimbing : Dra. Ida Widaningrum, M.Kom.
7. Konsultasi :

| NO | TANGGAL | URAIAN | TANDA TANGAN |
|----|---------------|---|---|
| 1. | 14 April 2020 | Pembahasan Dataset yang digunakan |  |
| 2. | 5 Juni 2020 | Demonstrasi Aplikasi |  |
| 3. | 22 Juli 2020 | Revisi Use Case, Flowchart sistem, Sequence diagram |  |
| 4. | 24 Juli 2020 | Revisi penulisan, abstrak, sitasi teori |  |
| 5. | 30 Juli 2020 | Ace Sidang | |






Ponorogo, 30 Juli 2020
Pembimbing,



(Dra. Ida Widaningrum, M.Kom)
NIK. 19660417 201101 13

**BERITA ACARA
BIMBINGAN SKRIPSI**

1. Nama : Wahidun Ni'am Fithriono Asyrofi
2. NIM : 16532569
3. Program Studi : Teknik Informatika
4. Fakultas : Teknik
8. Judul Skripsi : Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual
5. Dosen Pembimbing : Dyah Mustikasari, S.T. M.Eng.
6. Konsultasi :

| NO | TANGGAL | URAIAN | TANDA TANGAN |
|----|---------------|---|---|
| 1. | 14 April 2020 | Pembahasan dataset yang digunakan |  |
| 2. | 5 Juni 2020 | Demonstrasi Aplikasi |  |
| 3. | 22 Juli 2020 | Revisi Use Case, Flowchart, Sequence diagram, Pengujian |  |
| 4. | 24 Juli 2020 | Analisa dan revisi penulisan, nomor gambar, abstrak dan |  |
| 5. | 28 Jul 2020 | Sitasi Teori Ace Caday |  |

Ponorogo, 30 Juli 2020
Pembimbing,



(Dyah Mustikasari, S.T. M.Eng)
NIK. 19871007 201609 13



UNIVERSITAS MUHAMMADIYAH PONOROGO
UNIT PELAKSANA TEKNIS PERPUSTAKAAN
Jalan Budi Utomo 10 Ponorogo 63471 Jawa Timur Indonesia
Telp (0352) 481124, 487662 Fax (0352) 461796, *Website: library.umpo.ac.id*
TERAKREDITASI A
(SK Nomor 00012/ LAP.PT/ I.2017)

SURAT KETERANGAN
HASIL PEMERIKSAAN PLAGIASI SKRIPSI MAHASISWA
UNIVERSITAS MUHAMMADIYAH PONOROGO

Dengan ini kami nyatakan bahwa skripsi dengan rincian sebagai berikut:

Nama : Wahidun Ni`am Fithriono Asyrofi

NIM : 16532569

Prodi : Teknik Informatika

Judul : Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual

Dosen pembimbing :

1. Dra. Ida Widaningrum, M.Kom. email : iwidaningrum.as@gmail.com

2. Dyah Mustikasari, S.T. M.Eng. email : dyah.mustikasari@gmail.com

Telah dilakukan check plagiasi di UPT. Perpustakaan Universitas Muhammadiyah Ponorogo dengan prosentase plagiasi sebesar 10 %

Demikian keterangan ini dibuat untuk digunakan sebagaimana mestinya.

Ponorogo, 30 Juli 2020
Pemeriksa



(Mohamad Ulil Albab, SIP)
NIK.1989092720150322

Nb: Dosen pembimbing dimohon untuk mengecek kembali keaslian soft file karya ilmiah yang telah dicek di perpustakaan.



UNIVERSITAS MUHAMMADIYAH PONOROGO
UNIT PELAKSANA TEKNIS PERPUSTAKAAN
Jalan Budi Utomo No. 10 Ponorogo 63471 Jawa Timur Indonesia
Telp. (0352) 481124, Fax (0352) 461796, e-mail : lib@umpo.ac.id
website : www.library.umpo.ac.id

SURAT KETERANGAN
HASIL PEMERIKSAAN ANTI PLAGIASI ARTIKEL ILMIAH MAHASISWA
UNIVERSITAS MUHAMMADIYAH PONOROGO

Telah di periksa, artikel ilmiah dengan rincian sebagai berikut :

Nama : Wahidun Ni'am Fithriono Asyrofi
Judul : Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual
Fakultas / Prodi : Fakultas Teknik / Teknik Informatika

Dosen pembimbing :

1. Dra. Ida Widaningrum, M.Kom. email : iwidaningrum.as@gmail.com
2. Dyah Mustikasari, S.T. M.Eng. email : dyah.mustikasari@gmail.com

Tingkat kesamaan pada artikel sebesar 2%

Menggunakan aplikasi anti-plagiasi *Turnitin*. Demikian, atas perhatiannya di ucapkan terima kasih.

Ponorogo, 11 September 2020

Pemeriksa



(Mohamad Ulil Albab, SIP)

NIK.1989092720150322

Nb: Dosen pembimbing dimohon untuk mengecek kembali keaslian soft file karya ilmiah yang telah dicek di perpustakaan.

HALAMAN MOTTO

“Barangsiapa hari ini lebih baik daripada hari kemarin, maka ia adalah orang yang beruntung. Barangsiapa hari ini sama dengan hari kemarin, maka ia adalah orang yang merugi. Dan barangsiapa hari ini lebih buruk daripada hari kemarin, maka ia adalah orang yang terlaknat.”

-Ali bin Abi Thalib Radhiyallahu 'anhu

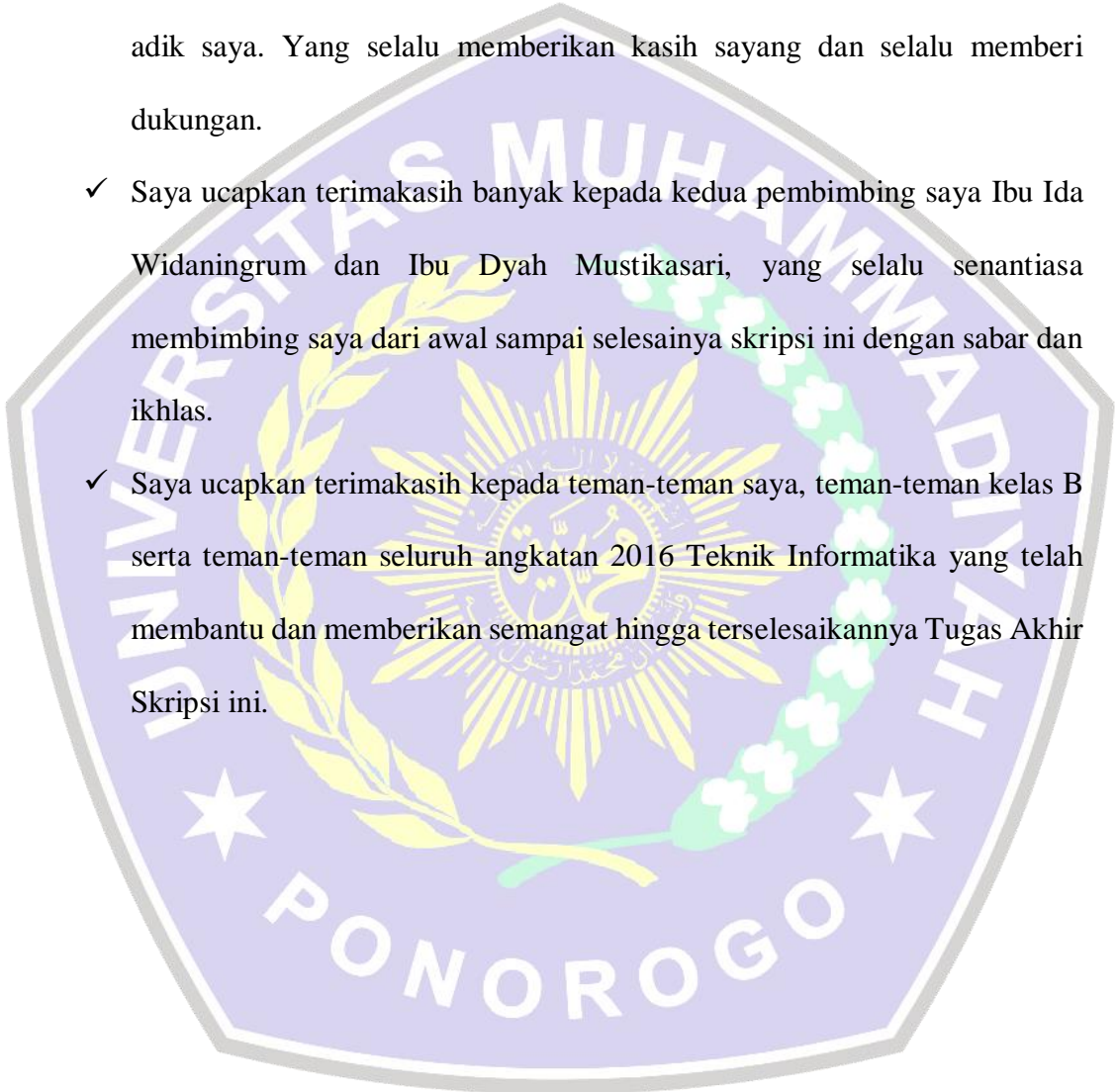
It's okay to lose people, but never lose yourself.

- Eminem



HALAMAN PERSEMBAHAN

- ✓ Yang pertama saya ucapkan puja dan puji syukur kehadiran Allah SWT yang telah memberikan kesehatan, kemudahan, dan kelancaran sehingga Tugas Akhir Skripsi ini bisa selesai tepat pada waktunya.
- ✓ Saya ucapkan banyak terimakasih kepada keluarga saya, yaitu orangtua dan adik saya. Yang selalu memberikan kasih sayang dan selalu memberi dukungan.
- ✓ Saya ucapkan terimakasih banyak kepada kedua pembimbing saya Ibu Ida Widaningrum dan Ibu Dyah Mustikasari, yang selalu senantiasa membimbing saya dari awal sampai selesainya skripsi ini dengan sabar dan ikhlas.
- ✓ Saya ucapkan terimakasih kepada teman-teman saya, teman-teman kelas B serta teman-teman seluruh angkatan 2016 Teknik Informatika yang telah membantu dan memberikan semangat hingga terselesaikannya Tugas Akhir Skripsi ini.



KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan rahmat, taufik serta hidayahnya sehingga penulis dapat menyelesaikan Tugas Akhir Skripsi ini dengan tepat waktu. Judul dari Tugas Akhir ini adalah “Evaluasi Pengukuran Jarak Pada Algoritma Winnowing Untuk Mendeteksi Plagiasi Bilingual” .

Tugas Akhir Skripsi ini disusun sebagai syarat Kelulusan pada Program Studi Teknik Informatika jenjang Strata Satu (S1) Fakultas Teknik Universitas Muhammadiyah Ponorogo.

Pada penyusunan Tugas Akhir Skripsi ini, telah banyak melibatkan beberapa pihak yang sangat membantu dan mendukung dalam berbagai hal. Maka dari itu, penulis ucapkan banyak terima kasih kepada :

1. Bapak Dr. Ir. Aliyadi, MM., M.Kom. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Ponorogo.
2. Ibu Dyah Mustikasari, S.T., M.Eng selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Ponorogo dan selaku Dosen Pembimbing II .
3. Ibu Dra. Ida Widaningrum, M.Kom selaku Dosen Pembimbing I
4. Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Ponorogo yang dengan kesabaran serta ketulusan hati memberikan bekal ilmu selama perkuliahan di Universitas Muhammadiyah Ponorogo.
5. Orang Tua tercinta yang telah memberikan doa dan dukungan kepada penulis.
6. Teman-teman Teknik Informatika angkatan 2016 Universitas Muhammadiyah Ponorogo serta semua pihak yang telah banyak membantu dalam penyusunan skripsi ini.

Semoga Allah SWT memberikan balasan yang jauh lebih sempurna. Penulis telah berusaha semaksimal mungkin dalam menyelesaikan Tugas Akhir Skripsi ini,akan tetapi penulis sadar bahwa Tugas Akhir Skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun senantiasa penulis harapkan. Semoga Tugas Akhir Skripsi ini bermanfaat bagi kita semua. Amin.

Ponorogo, Agustus 2020

Wahidun Ni'am Fithriono Asyrofi

EVALUASI PENGUKURAN JARAK PADA ALGORITMA WINNOWER UNTUK MENDETEKSI PLAGIASI BILINGUAL

Wahidun Ni'am Fithriono Asyofi ¹⁾, Ida Widaningrum ²⁾, Dyah Mustikasari ³⁾

^{1,2,3)} Program Studi Teknik Informatika, Fakultas Teknik,

Universitas Muhammadiyah Ponorogo

e-mail: iam.wahidn@gmail.com ¹⁾, iwidaningrum.as@gmail.com ²⁾,
dyah.mustikasari@gmail.com ³⁾

ABSTRAK

Pembandingan teks karya ilmiah bisa dengan cara menghitung kata atau kalimat yang sama dan ada pada karya ilmiah lainnya. Namun untuk cara ini membutuhkan pemrosesan yang panjang mengingat semua katanya akan dibandingkan serta dipengaruhi banyaknya teks karya ilmiah yang dibandingkan. Cara lain untuk membandingkan teks karya ilmiah adalah dengan menjadikannya sebagai *fingerprint*. Metode *Fingerprinting* Algoritma Winnower merupakan salah satu metode yang mengubah dokumen menjadi kumpulan hash yang mewakili teks karya ilmiah dan mempersingkat penulisan karena tiap kata atau kalimat diperingkas dengan hashing. Penghitungan jarak kesamaan menggunakan beberapa metode pengukuran. Metode tersebut adalah penghitungan jarak yang akan diterapkan yaitu Cosine, Jaccard, Dice, Euclidean, Manhattan, Minkowski, Mahalanobis dan Weighted. Tiap rumus berbeda hasil pengukuran satu dan lainnya yang merupakan karakteristik dari masing-masing metode pengukuran. Dari hasil pengujian menunjukkan bahwa akurasi pengukuran terhadap nilai aproksimasi yang ditentukan menunjukkan metode pengukuran Cosine Similarity merupakan metode yang lebih baik dibanding ke 7 metode pengukuran lainnya pada penelitian ini. Ini dibuktikan dengan rerata selisih 1.4725 dan cukup stabil dengan standar deviasi 1.4823. Pada metode Dice Similarity memiliki selisih rerata 2.8475 dan standar deviasi 1.8393. Metode Jaccard Similarity selisih reratanya 12.689 dan standar deviasi 3.5664. Akurasinya masuk dalam kategori menengah dan stabil pada beberapa pengujian. Pada metode Manhattan, Minkowski, Euclidean, Mahalanobis dan Weighted selisih reratanya lebih besar dari 21 dan standar deviasinya lebih besar dari 16 yang menunjukkan bahwa akurasinya lebih rendah dan kurang stabil pada beberapa pengujian.

Kata Kunci: Algoritma Winnower, *Cosine Similarity*, *Jaccard Similarity*, *Dice Similarity*, *Euclidean Distance*, *Manhattan Distance*, *Minkowski Distance*, *Mahalanobis Distance*, *Weighted Distance*

DAFTAR ISI

| | |
|---|------|
| HALAMAN JUDUL | i |
| HALAMAN PENGESAHAN | ii |
| PERNYATAAN ORISINALITAS SKRIPSI | iii |
| HALAMAN BERITA ACARA UJIAN | iv |
| BERITA ACARA BIMBINGAN | v |
| SURAT KETERANGAN HASIL PLAGIASI SKRIPSI | vii |
| SURAT KETERANGAN HASIL PLAGIASI ARTIKEL | viii |
| HALAMAN MOTTO | ix |
| HALAMAN PERSEMBAHAN | x |
| KATA PENGANTAR | xi |
| ABSTRAK | xii |
| DAFTAR ISI | xiii |
| DAFTAR GAMBAR | xv |
| DAFTAR TABEL | xvi |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang | 1 |
| 1.2. Rumusan Masalah | 2 |
| 1.3. Tujuan Penelitian | 2 |
| 1.4. Manfaat Penelitian | 3 |
| 1.5. Batasan Masalah | 3 |
| BAB II TINJAUAN PUSTAKA | 4 |
| 2.1. Algoritma Winnowing | 4 |
| 2.2. N-Gram | 5 |
| 2.3. Rolling Hash | 5 |
| 2.4. Cosine Similarity | 6 |
| 2.5. Jaccard Similarity | 6 |
| 2.6. Dice Similarity | 7 |
| 2.7. Euclidean Distance | 7 |
| 2.8. Manhattan Distance | 7 |
| 2.9. Minkowski Distance | 8 |
| 2.10. Mahalanobis Distance | 8 |
| 2.11. Weighted Distance | 9 |

| | | |
|--|--|-----------|
| 2.12. | Bilingual | 9 |
| 2.13. | Google Translate API | 9 |
| 2.14. | Python | 11 |
| 2.15. | Penelitian Terdahulu | 11 |
| BAB III METODE PENELITIAN | | 14 |
| 3.1. | Tahap Penelitian | 14 |
| 3.2. | Alat dan Bahan Penelitian | 16 |
| 3.2.1. | Alat Penelitian | 16 |
| 3.2.2. | Bahan Penelitian | 16 |
| 3.3. | Metode Penelitian | 16 |
| 3.3.1. | Metode Pengumpulan Data | 16 |
| 3.3.2. | Metode Pengembangan Perangkat Lunak | 17 |
| 3.4. | Model Flowchart Sistem | 19 |
| BAB IV PEMBAHASAN | | 23 |
| 4.1. | Analisis Sistem | 23 |
| 4.1.1. | Penggambaran Sistem | 23 |
| 4.1.2. | Struktur Database | 24 |
| 4.1.3. | Use Case Diagram | 26 |
| 4.1.4. | Sequence Diagram Kelola Dataset | 27 |
| 4.1.5. | Sequence Diagram Kelola Dokumen Uji | 28 |
| 4.2. | Pembahasan Alur Sistem | 29 |
| 4.3.1. | Autentikasi | 29 |
| 4.3.2. | Mengelola Dataset | 30 |
| 4.3.3. | Mengelola Dokumen Uji | 39 |
| 4.3. | Analisis Hasil Pengukuran Kesamaan dan Jarak Perbedaan | 46 |
| 4.4.1. | Hasil Data Uji | 48 |
| BAB V PENUTUP | | 55 |
| 5.1. | Kesimpulan | 55 |
| 5.2. | Saran | 56 |
| DAFTAR PUSTAKA | | 57 |
| LAMPIRAN | | 59 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.13.1 Google API Translate..... | 10 |
| Gambar 3.1.1 Desain Penelitian | 14 |
| Gambar 3.3.1 Model Pengembangan Spiral..... | 17 |
| Gambar 3.4.1 Flowchart Sistem | 19 |
| Gambar 4.1.1 Tabel Document | 24 |
| Gambar 4.1.2 Tabel Similarity | 25 |
| Gambar 4.1.3 Tabel User | 25 |
| Gambar 4.1.4 Use Case Diagram | 26 |
| Gambar 4.1.5 Sequence Diagram Kelola Dataset | 27 |
| Gambar 4.1.6 Sequence Diagram Kelola Dokumen..... | 28 |
| Gambar 4.2.1 Form Login dan Register | 29 |
| Gambar 4.2.2 Sidebar Menu..... | 30 |
| Gambar 4.2.3 Form Upload Dataset | 31 |
| Gambar 4.2.4 Sanitize Text..... | 32 |
| Gambar 4.2.5 Hasil Sanitize Text..... | 32 |
| Gambar 4.2.6 N-Gram | 33 |
| Gambar 4.2.7 Hasil N-Gram panjang 7 | 33 |
| Gambar 4.2.8 Rolling Hash Function | 34 |
| Gambar 4.2.9 Hasil Rolling Hash..... | 35 |
| Gambar 4.2.10 Rangkaian Window dari Hash..... | 35 |
| Gambar 4.2.11 <i>Fingerprinting</i> | 36 |
| Gambar 4.2.12 Hasil <i>Fingerprinting</i> | 36 |
| Gambar 4.2.13 Halaman Daftar Dataset | 37 |
| Gambar 4.2.14 Form Edit Dataset | 37 |
| Gambar 4.2.15 Halaman Detail Dataset..... | 38 |
| Gambar 4.2.16 Tampilan debug Dataset..... | 38 |
| Gambar 4.2.17 Halaman Upload Dokumen | 39 |
| Gambar 4.2.18 Mengambil <i>fingerprint</i> dokumen..... | 41 |
| Gambar 4.2.19 Mengambil kumpulan <i>fingerprint</i> dataset..... | 41 |
| Gambar 4.2.20 Similarity and Distance Measuring..... | 42 |
| Gambar 4.2.21 Halaman Daftar Dokumen..... | 43 |
| Gambar 4.2.22 Halaman Detail Dokumen | 43 |
| Gambar 4.2.23 Tampilan Menu Detail | 44 |
| Gambar 4.2.24 Tampilan Tabel Hasil Pengukuran | 44 |
| Gambar 4.2.25 Grafik Kesamaan | 45 |
| Gambar 4.2.26 Grafik Jarak Perbedaan | 45 |
| Gambar 4.3.1 Hasil Pengukuran..... | 47 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 4.1 Hasil Pengujian Test 1..... | 49 |
| Tabel 4.2 Selisih hasil Test 1 dengan nilai perkiraan | 49 |
| Tabel 4.3 Hasil Pengujian Test 2..... | 49 |
| Tabel 4.4 Selisih hasil Test 2 dengan nilai perkiraan | 50 |
| Tabel 4.5 Hasil Pengujian Test 3..... | 50 |
| Tabel 4.6 Selisih hasil Test 3 dengan nilai perkiraan | 51 |
| Tabel 4.7 Hasil Pengujian Test 4..... | 51 |
| Tabel 4.8 Selisih hasil Test 4 dengan nilai perkiraan | 52 |
| Tabel 4.9 Hasil Pengujian Test 5..... | 52 |
| Tabel 4.10 Selisih hasil Test 5 dengan nilai perkiraan | 53 |
| Tabel 4.11 Rata-rata selisih tiap metode pengukuran..... | 54 |

