

**APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN
METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB***

Skripsi



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2024

**APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN
METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB***

Skripsi

**Diajukan sebagai salah satu syarat untuk kelengkapan gelar kesarjanaan pada
Program Studi Teknik Informatika
Jenjang Pendidikan Strata 1**



Jimmy Halim

2020100023

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS BUDDHI DHARMA

TANGERANG

2024

LEMBAR PERSEMBAHAN

Ἡ ἐκπαίδευση τροφοδοτεῖ τὴ φωτιᾶ, ὄχι γεμίζει τὸ σκάφος

“Pendidikan itu mengobarkan api, bukan mengisi bejana”

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa, SKRIPSI ini

kupersembahkan untuk:

1. Bapak Lian Siong dan Ibu Bie Jen tercinta yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
2. Kakak-kakakku yang telah memberikan dukungan semangat serta dorongan yang senantiasa diberikan.
3. Teman-teman kelompok belajar yang selalu berjuang bersama dalam sedih maupun senang.
4. Bu Desi yang penuh dengan kesabaran dalam memberikan bimbingan, saran maupun arahan, serta masukan yang sangat berarti di dalam proses-proses yang melelahkan dalam penyusunan skripsi ini, terima kasih atas segala bimbingan dan ilmu yang telah diberikan.
5. Rekan-rekan SEMA FST yang telah bersedia untuk berkerja-sama dalam suka dan duka, terima kasih atas segala waktu dan kebersamaannya.
6. Universitas yang telah berjasa untuk menjaga dari awal hingga akhir, serta memberikan ilmu pengetahuan yang dibutuhkan.

UNIVERSITAS BUDDHI DHARMA

LEMBARAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan dibawah ini.

NIM : 20201000023
Nama : Jimmy Halim
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : Database Development

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapat gelar akademik Sarjana atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar Pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak benaran dalam pernyataan hari ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 29 Juli 2024

Yang membuat pernyataan,



Jimmy Halim

20201000023

UNIVERSITAS BUDDHI DHARMA

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini.

NIM : 20201000023
Nama : Jimmy Halim
Jenjang Studi : Strata I
Program Studi : Teknik Informatika
Peminatan : Database Development

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul: “Aplikasi Deteksi Plagiarisme Dokumen dengan Metode *TF-IDF* dan *Cosine Similarity* Berbasis *Web*”, beserta alat yang diperlukan (apabila ada).

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 29 Juli 2024

Yang membuat pernyataan,



Jimmy Halim

20201000023

UNIVERSITAS BUDDHI DHARMA

LEMBAR PENGESAHAN PEMBIMBING
APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN
METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB*

Dibuat oleh:

NIM : 20201000023

Nama : Jimmy Halim

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

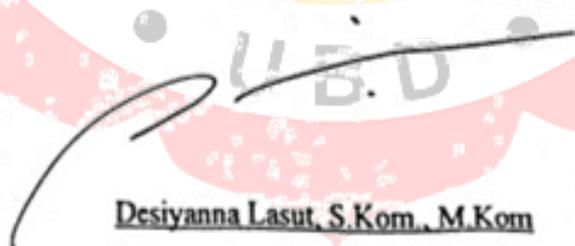
Peminatan Database Development

Tahun Akademik 2023/2024

Disahkan oleh,

Tangerang, 21 Juli 2024

Pembimbing


Desiyanna Lasut, S.Kom., M.Kom

NIDN : 0402128601

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN SKRIPSI
APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN
METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB*

Dibuat Oleh:

NIM : 20201000023

Nama : Jimmy Halim

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Peminatan Database Development

Tahun Akademik 2023/2024

Disahkan oleh,

Tangerang, 29 Juli 2024

Dekan

Ketua Program Studi


Dr. Yakub, M.Kom., M.M.

NIDN: 0304056901


Hartana Wijaya, M.Kom.

NIDN: 0412058102

LEMBAR PENGESAHAN TIM PENGUJI

Nama : Jimmy Halim
NIM : 20201000023
Fakultas : Sains dan Teknologi
Judul Skripsi : APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN
METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS
WEB

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Senin, 29 Juli
2024

Nama Penguji : Tanda Tangan :
Ketua Sidang : **Rino, M.Kom.**
NIDN : 0420058502
Penguji I : **Susanto Hariyanto, M.Kom.**
NIDN : 0428128601
Penguji II : **Desiyanna Lasut, M.Kom.**
NIDN : 0402128601

Mengetahui,

Dekan Fakultas Sains dan Teknologi

Dr. Yakub, M.Kom., M.M.

NIDN : 0304056901

KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul Judul APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB*. Tujuan utama dari pembuatan Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program Pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P. sebagai Rektor Universitas Buddhi Dharma.
2. Bapak Dr. Yakub, M.Kom., M.M. Dekan Fakultas Sains dan Teknologi.
3. Bapak Hartana Wijaya, S.Kom., M.Kom. sebagai Ketua Program Studi Teknik Informatika.
4. Ibu Desiyanna Lasut, S.Kom., M.Kom sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penelitian Skripsi ini.
5. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
6. Teman-teman yang selalu membantu dan memberikan semangat.

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penelitian ini. Penulis menyadari bahwa penelitian Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penelitian di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 29 Juli 2024

Penulis

APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB*

114 Halaman + xviii/ 45 tabel / 69 gambar / 7 lampiran

ABSTRAK

Dalam era teknologi dan informasi yang sangat berkembang pesat akhir-akhir ini, hal tersebut menyebabkan akses yang mudah terhadap informasi yang mempermudah proses belajar di dunia pendidikan, namun kemudahan tersebut juga memicu tindakan-tindakan plagiarisme yang dimana menjadi ancaman serius bagi ilmu pengetahuan. Plagiarisme adalah sebuah tindakan mencuri atau mengambil karya orang lain tanpa memberikan atribusi yang layak atau dapat dibilang tanpa mencitasi orang tersebut. Oleh karena itu, maka dikembangkanlah sebuah aplikasi yang dapat mengatasi masalah tersebut, yaitu aplikasi pendeteksi plagiarisme yang menggunakan metode algoritma *TF-IDF* (*Term Frequency-Inverse Document Frequency*) dan *Cosine Similarity*. *TF-IDF* dan *Cosine Similarity* akan diimplementasikan ke dalam aplikasi untuk melakukan proses perhitungan yang dimana akhirnya akan memberikan hasil yang berupa persentase dari perhitungan yang telah dilakukan. Aplikasi plagiarisme ini dirancang untuk mendeteksi kesamaan antara dokumen dalam database dan dokumen user. Proses-proses yang terjadi di dalam aplikasi meliputi proses *preprocessing*, perhitungan *TF-IDF*, dan perhitungan *Cosine Similarity*. Hasil pengujian yang dilakukan dapat dikatakan konsisten karena hasil pengujian manual dan aplikasi menunjukkan hasil persentase yaitu 4% dan 4,34%. Aplikasi akan juga melewati proses kuesioner yang dilakukan, agar user dapat memberikan pendapat-pendapat mereka apakah aplikasi akan layak atau tidak, setelah dilakukannya proses kuesioner didapatkan hasil secara garis besar user merasa puas dan sangat puas dengan aplikasi yang dibuat.

Kata Kunci : Teknik Informatika akademik, *Database*, *Plagiarism*, *TF-IDF*, *Cosine Similarity*

APLIKASI DETEKSI PLAGIARISME DOKUMEN DENGAN METODE *TF-IDF* DAN *COSINE SIMILARITY* BERBASIS *WEB*

114 Pages + xviii / 45 Table / 69 Picture / 7 Appendices

ABSTRACT

*In the era of technology and information which is developing very rapidly recently, this has resulted in easy access to information which makes the learning process easier in the world of education, but this ease also triggers acts of plagiarism which is a serious threat to science. Plagiarism is an act of stealing or taking someone else's work without giving proper attribution or you could say without citing that person. Therefore, an application was developed that can overcome this problem, namely a plagiarism detection application that uses the *TF-IDF* (Term Frequency-Inverse Document Frequency) and Cosine Similarity algorithm methods. *tfi-idf* and Cosine Similarity will be implemented into the application to carry out the calculation process which will ultimately provide results in the form of a percentage of the calculations that have been carried out. This plagiarism application is designed to detect similarities between documents in the database and user documents. The processes that occur in the application include preprocessing processes, *TF-IDF* calculations, and Cosine Similarity calculations. The results of the tests carried out can be said to be consistent because the results of manual and application tests show percentage results of 4% and 4.34%. The application will also go through a questionnaire process, so that users can provide their opinions on whether the application will be feasible or not. After carrying out the questionnaire process, the results obtained are that the user is generally satisfied and very satisfied with the application created.*

*Keywords: Academic Informatics Engineering, Database, Plagiarism, *TF-IDF*, Cosine Similarity.*

DAFTAR ISI

LEMBARAN JUDUL DALAM

LEMBARAN PERNYATAAN KEASLIAN SKRIPSI

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

LEMBAR PENGESAHAN PEMBIMBING

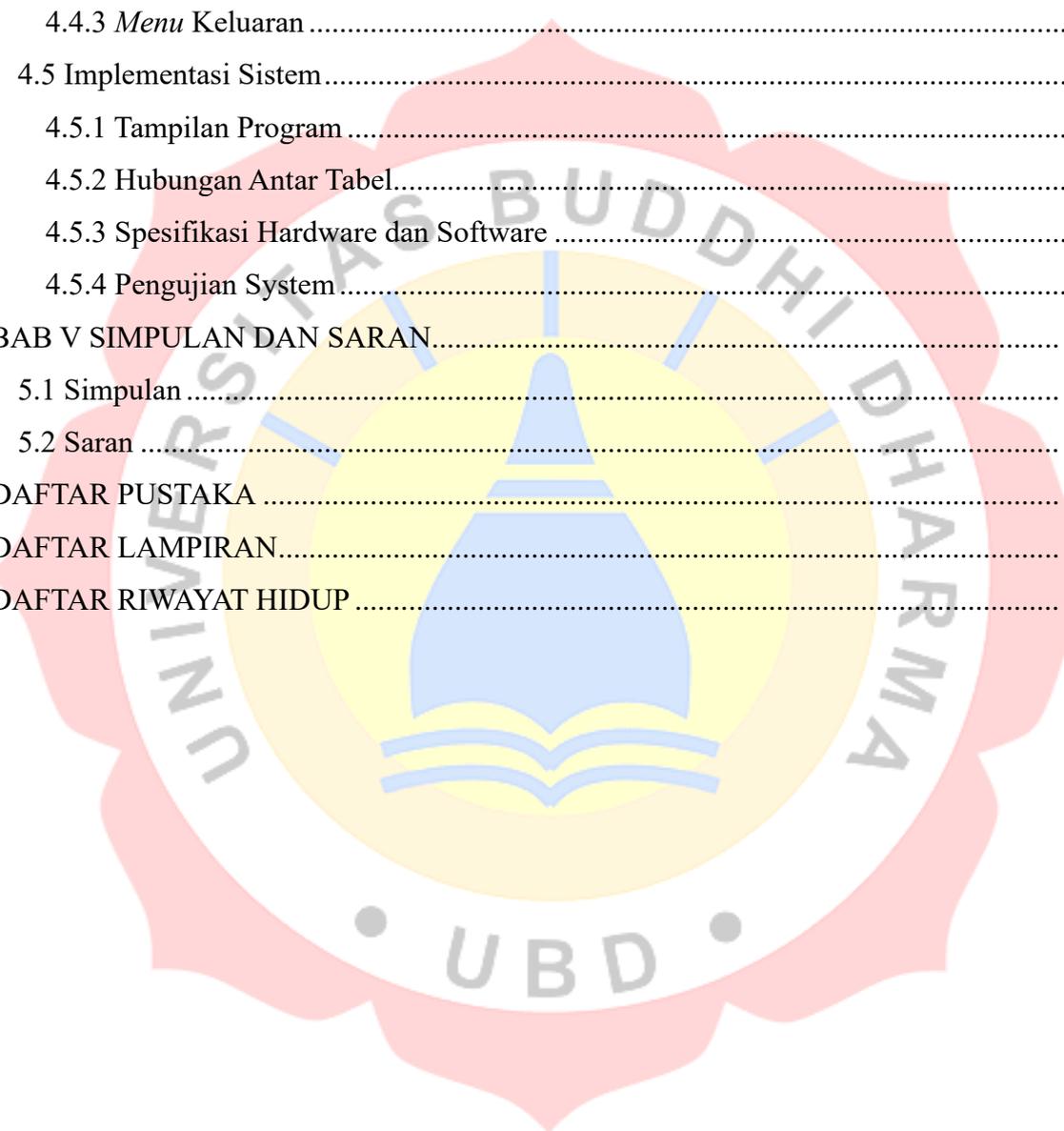
LEMBAR PENGESAHAN SKRIPSI

LEMBAR PENGESAHAN TIM PENGUJI

KATA PENGANTAR	vii
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah	3
1.2.1 Dalam dunia pendidikan masyarakat	3
1.2.2 Dalam dunia Ilmu Pengetahuan dan Teknologi (IPTEK)	3
1.3 Ruang Lingkup	3
1.4 Tujuan dan Manfaat	4
1.4.1 Tujuan	4
1.4.2 Manfaat	4
1.5 Teknik Pengumpulan Data	4
1.5.1 Studi Pustaka	4
1.5.2 `Kuesioner	5
1.6 Sistematisan Penelitian	5
BAB II LANDASAN TEORI	7
2.1 Teori Umum	7
2.1.1 Data	7
2.1.2 Informasi	8
2.2 Teori Khusus	9
2.2.1 Plagiarisme	9

2.2.2 <i>Text Mining</i>	10
2.2.3 <i>NLP (Natural language processing)</i>	12
2.2.4 <i>TF-IDF</i>	13
2.2.5 <i>Cosine Similarity</i>	13
2.2.6 <i>Waterfall</i>	14
2.3 Teori Analisa dan Perancangan	14
2.3.1 Teori Metodologi Penelitian.....	14
2.3.2 Teori Pemodelan Proses	19
2.3.3 Teori Aplikasi	25
2.4 Teori pengujian	29
2.4.1 <i>Blackbox Testing</i>	29
2.5 Tinjauan Studi.....	30
2.6 Kerangka Pemikiran	54
BAB III METODOLOGI PENELITIAN	55
3.1 <i>Data Flow Diagram (DFD)</i>	55
3.2 Analisa	55
3.2.1 Masukan	55
3.2.2 Proses	56
3.2.3 Keluaran	56
3.2.4 Data	56
3.3 Analisa Masalah.....	57
3.4 Identifikasi kebutuhan sistem	57
3.5 Metode <i>TF-IDF</i> dan <i>Cosine Similarity</i>	57
3.6 Requirement Elicitation	62
3.7 Jadwal Penelitian	62
BAB IV PENGUJIAN DAN IMPLEMENTASI	64
4.1 Prosedur Sistem Usulan	64
4.2 Rancangan Sistem Usulan	64
4.2.1 Bagan Berjenjang (<i>Hierarchy Charty</i>).....	64
4.2.2 Konteks Diagram (<i>Top Level</i>)	65
4.2.3 <i>Diagram Overview (Level 0)</i>	66
4.2.4 Diagram Rinci (<i>Level 1</i>).....	67
4.3 Rancangan Database	70
4.3.1 <i>Entity Relationship Diagram (ERD)</i>	70

4.3.2 Logical Record Structure (LRS)	71
4.3.3 Normalisasi Data (<i>Normal 3</i>)	72
4.3.4 Struktur File (Spesifikasi Basis Data)	72
4.4 Rancangan Tampilan Program	74
4.4.1 <i>Menu</i> Utama	74
4.4.2 <i>Menu</i> Masukan	75
4.4.3 <i>Menu</i> Keluaran	83
4.5 Implementasi Sistem	85
4.5.1 Tampilan Program	85
4.5.2 Hubungan Antar Tabel	92
4.5.3 Spesifikasi Hardware dan Software	93
4.5.4 Pengujian System	93
BAB V SIMPULAN DAN SARAN	113
5.1 Simpulan	113
5.2 Saran	113
DAFTAR PUSTAKA	114
DAFTAR LAMPIRAN	118
DAFTAR RIWAYAT HIDUP	125



DAFTAR GAMBAR

Gambar 2.1 Metode Waterfall	14
Gambar 2.2 Feature Selection	18
Gambar 2.3 <i>Flow Direction Symbols</i>	19
Gambar 2.4 <i>Processing symbols</i>	20
Gambar 2.5 <i>Processing Symbols</i>	20
Gambar 2.6 <i>Input/Output Symbols</i>	21
Gambar 2.7 Input/Output Symbols	21
Gambar 2.8 Entity Relationship Diagram (ERD)	22
Gambar 2.9 Diagram Berjenjang	23
Gambar 2.10 Diagram Konteks	23
Gambar 2.11 Data Flow Daigram Level 0	24
Gambar 2.12 Data Flow Diagram Level 1	24
Gambar 2.13 Logical record Structure (LRS).....	25
Gambar 2.14 Kerangka Pemikiran.....	54
Gambar 3.1 Data Flow Diagram	55
Gambar 4.1 Bagan Berjenjang	64
Gambar 4.2 Diagram Konteks	65
Gambar 4.3 Diagram Overview	66
Gambar 4.4 Diagram Rinci Login.....	67
Gambar 4.5 Diagram Rinci User yang sudah Login.....	68
Gambar 4.6 Diagram Rinci Admin	69
Gambar 4.7 Diagram Rinci User yang belum Login	69
Gambar 4.8 Entity Relationship Diagram.....	70
Gambar 4.9 Logical Record Structure	71
Gambar 4.10 Menu Utama dan Menu Masukan User yang belum Login	74
Gambar 4.11 Menu Masukan Register.....	75
Gambar 4.12 Menu Masukan Login	76
Gambar 4.13 Menu Masukan User yang sudah Login.....	76
Gambar 4.14 Menu Masukan Setting Account	77
Gambar 4.15 Menu Masukan Admin	78
Gambar 4.16 Menu Masukan Datasets	78
Gambar 4.17 Menu Masukan Edit Datasets.....	79

Gambar 4.18 Menu Masukan Admin User Login.....	80
Gambar 4.19 Menu Masukan Admin Edit User Login	80
Gambar 4.20 Menu Masukan Admin History User yang sudah Login.....	81
Gambar 4.21 Menu Masukan Admin Edit History User yang sudah Login	82
Gambar 4.22 Menu Masukan Admin History User yang belum Login	82
Gambar 4.23 Menu Hasil	83
Gambar 4. 24 Hasil Download.....	84
Gambar 4.25 Menu History User yang sudah Login	85
Gambar 4.26 Menu Utama dan Masukan untuk User yang belum Login	85
Gambar 4.27 Menu Masukan Register	86
Gambar 4.28 Menu Masukan Login	86
Gambar 4.29 Menu Masukan User yang sudah Login.....	86
Gambar 4.30 Menu Masukan Setting Account	87
Gambar 4.31 Menu Masukan Admin	87
Gambar 4.32 Menu Masukan Datasets	87
Gambar 4. 33 Menu Masukan Datasets	88
Gambar 4.34 Menu Masukan Edit Datasets.....	88
Gambar 4.35 Menu Masukan Admin User Login.....	88
Gambar 4.36 Menu Masukan Admin Edit User Login	89
Gambar 4.37 Menu Masukan Admin History User yang sudah Login.....	89
Gambar 4.38 Menu Masukkan Admin History User yang sudah Login.....	89
Gambar 4.39 Menu Masukan Admin History User yang belum Login	90
Gambar 4.40 Menu Hasil	90
Gambar 4.41 Menu Hasil	90
Gambar 4. 42 Hasil Download.....	91
Gambar 4.43 Menu History User yang sudah Login	92
Gambar 4.44 Hubungan Antar Tabel	92
Gambar 4.45 Perhitungan Pertanyaan 1	97
Gambar 4.46 Perhitungan Pertanyaan 2.....	98
Gambar 4.47 Perhitungan Pertanyaan 3	100
Gambar 4.48 Perhitungan Pertanyaan 4.....	101
Gambar 4.49 Perhitungan Pertanyaan 5.....	103
Gambar 4.50 Perhitungan Pertanyaan 6.....	104
Gambar 4.51 Perhitungan Pertanyaan 7.....	106

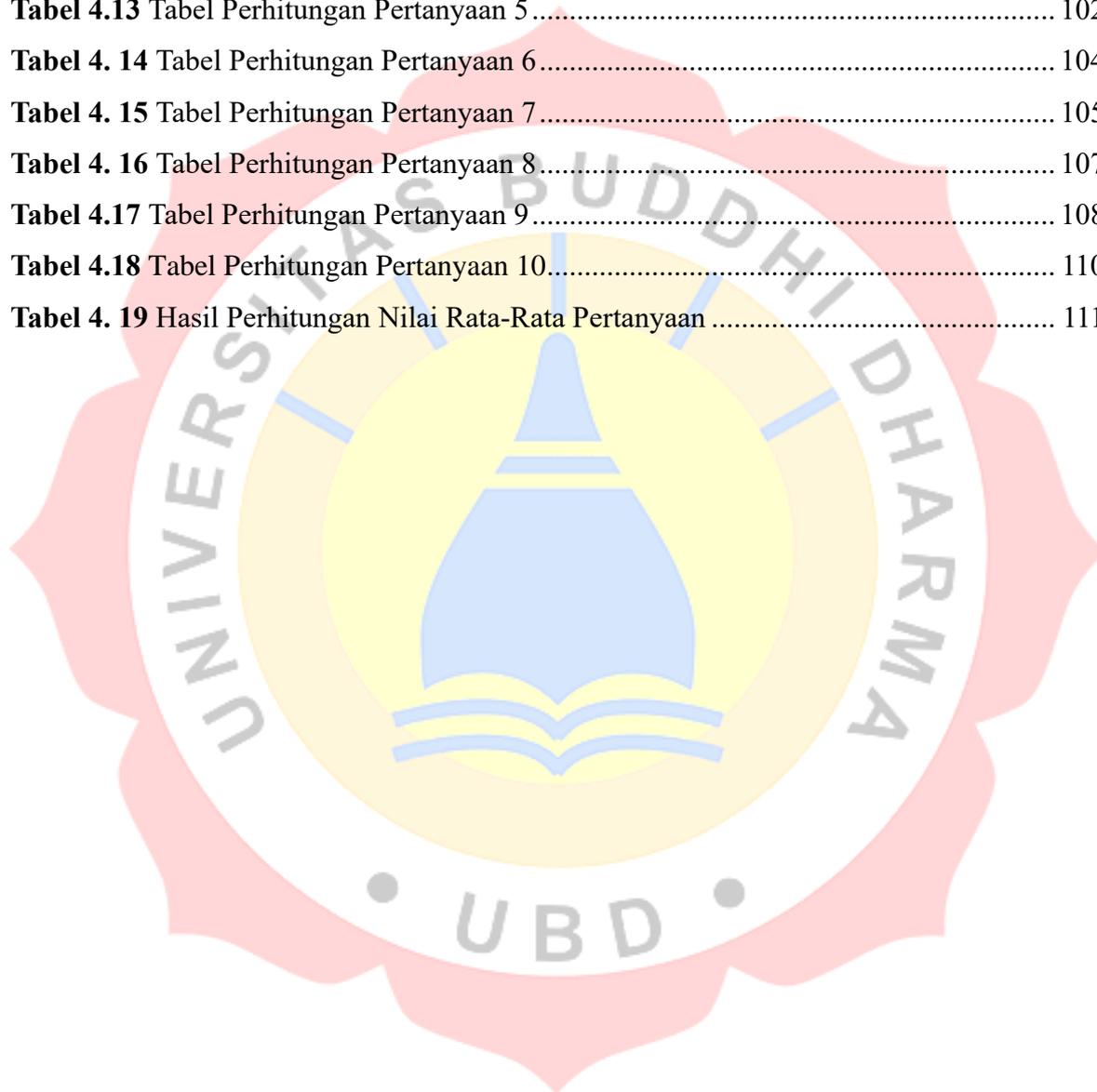
Gambar 4.52 Perhitungan Pertanyaan 8..... 107
Gambar 4.53 Perhitungan Pertanyaan 9..... 109
Gambar 4.54 Perhitungan Pertanyaan 10..... 110



DAFTAR TABEL

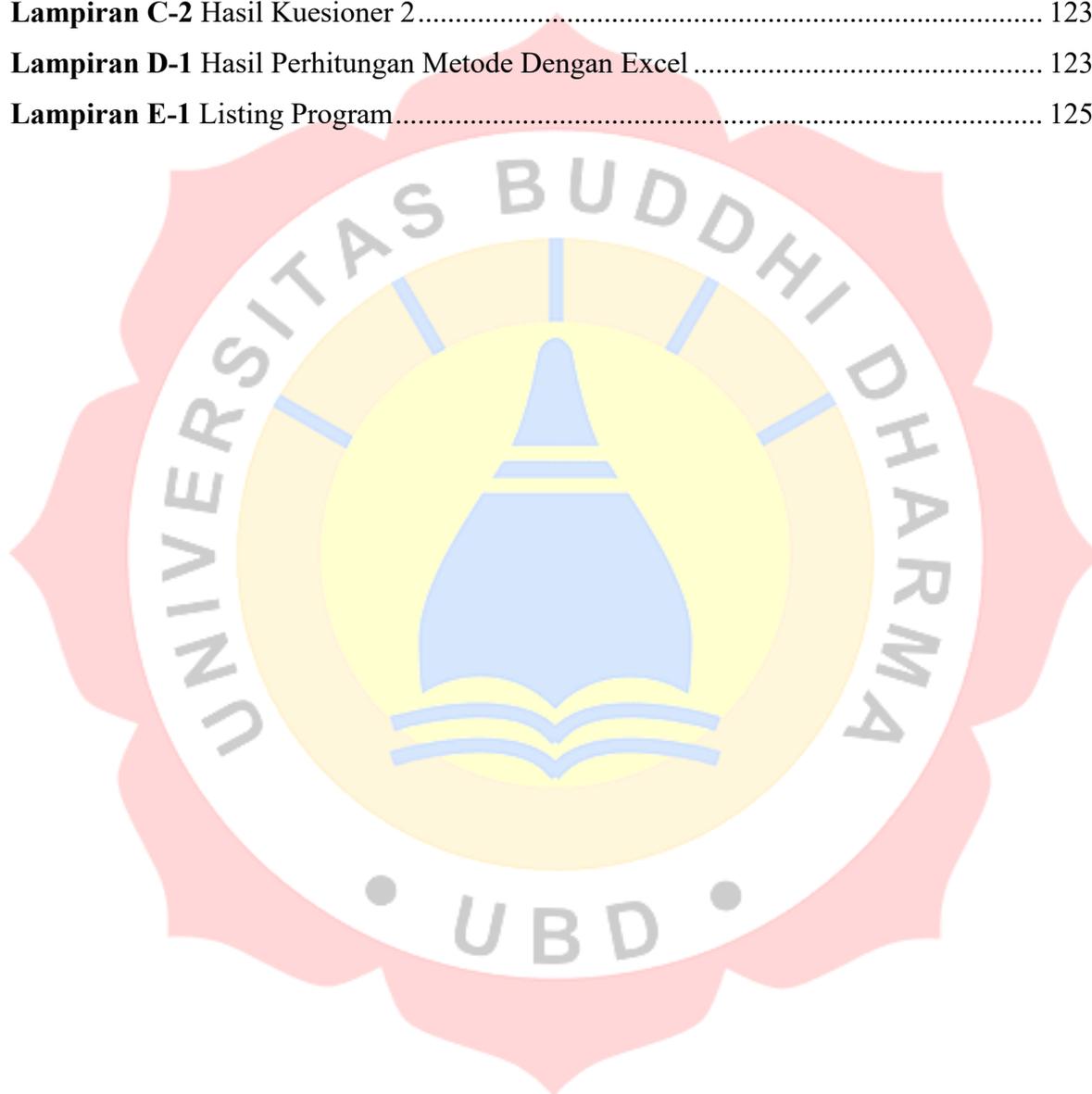
Tabel 2.1 Lowercasing	16
Tabel 2.2 Noise Removal	17
Tabel 2.3 Tabel Uji Black Box	29
Tabel 2.4 Jurnal ke-1	30
Tabel 2.5 Jurnal ke-2	31
Tabel 2.6 Jurnal ke-3	33
Tabel 2.7 Jurnal ke-4	36
Tabel 2.8 Jurnal ke-5	37
Tabel 2.9 Jurnal ke-6	38
Tabel 2.10 Jurnal ke-7	40
Tabel 2.11 Jurnal ke-8	41
Tabel 2.12 Jurnal ke-9	43
Tabel 2.13 Jurnal ke-10	44
Tabel 2.14 Jurnal ke-11	46
Tabel 2.15 Jurnal ke-12	47
Tabel 2.16 Jurnal ke-13	49
Tabel 2.17 Jurnal ke-14	50
Tabel 2.18 Jurnal ke-15	52
Tabel 3.1 Tabel tf	58
Tabel 3.2 Tabel df	59
Tabel 3.3 Tabel idf	59
Tabel 3.4 Tabel TF-IDF	60
Tabel 3.5 Tabel Hasil	61
Tabel 3.6 Requirement Elicitation	62
Tabel 3.7 Requirement Elicitation	62
Tabel 3.8 Tabel Jadwal Penelitian	62
Tabel 4.1 Struktur File User	73
Tabel 4.2 Struktur File User (login)	73
Tabel 4.3 Struktur File Login	73
Tabel 4.4 Struktur File Admin	74
Tabel 4.5 Tabel uji Black Box	93
Tabel 4.6 Skor Jawaban	95

Tabel 4.7 Skor Ideal	95
Tabel 4.8 Rating Scale	96
Tabel 4.9 Tabel Perhitungan Pertanyaan 1	96
Tabel 4.10 Tabel Perhitungan Pertanyaan 2	98
Tabel 4.11 Tabel Perhitungan Pertanyaan 3	99
Tabel 4.12 Tabel Perhitungan Pertanyaan 4	101
Tabel 4.13 Tabel Perhitungan Pertanyaan 5	102
Tabel 4.14 Tabel Perhitungan Pertanyaan 6	104
Tabel 4.15 Tabel Perhitungan Pertanyaan 7	105
Tabel 4.16 Tabel Perhitungan Pertanyaan 8	107
Tabel 4.17 Tabel Perhitungan Pertanyaan 9	108
Tabel 4.18 Tabel Perhitungan Pertanyaan 10	110
Tabel 4.19 Hasil Perhitungan Nilai Rata-Rata Pertanyaan	111



DAFTAR LAMPIRAN

Lampiran A-1 Kartu Bimbingan.....	119
Lampiran B-1 <i>Requirement Elicitation 1</i>	120
Lampiran B-2 <i>Requirement Elicitation 2</i>	121
Lampiran C-1 Hasil Kuesioner 1.....	122
Lampiran C-2 Hasil Kuesioner 2.....	123
Lampiran D-1 Hasil Perhitungan Metode Dengan Excel.....	123
Lampiran E-1 Listing Program.....	125



BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dengan perkembangan jaman teknologi dan informasi yang sangat pesat akhir-akhir ini, siapapun dapat dengan mudah menemukan dan mencari informasi yang diperlukan. Perkembangan dari teknologi dan informasi tersebut memberikan dampak yang sangat besar bagi manusia akhir-akhir ini, contohnya dunia pendidikan, yang dimana hal ini menyebabkan mahasiswa atau pelajar diberikan akses secara luas dalam mencari informasi, mulai dari buku yang ada di internet, jurnal yang disebar melalui *google*, tutorial dari *youtube*, dan lain-lain. Di dunia pendidikan yang didukung oleh teknologi ini, informasi-informasi dari *e-book*, artikel, dan jurnal baik yang berskala internasional atau nasional dengan reputasi yang bervariasi dapat dengan mudah diakses melalui internet (L. Hermawan & Ismiati, 2020, p. 94). Walaupun kemudahan ini diberikan, pasti ada Tindakan-tindakan yang memicu penyimpangan, seperti plagiarisme.

Plagiarisme adalah tindakan yang melanggar, tindakan-tindakan ini menjadi ancaman yang besar bagi ilmu pengetahuan. Pelaku dari plagiarisme biasanya mencuri karya-karya penulis lain ke dalam karya mereka tanpa mengutip referensi yang aslinya (Shadiqi, 2019, p. 30).

Semua organisasi ini sepakat bahwa kalau pelanggaran dalam penelitian ilmiah itu bisa terjadi pada dasarnya melalui tiga praktik yang dikutuk oleh para peneliti : Memalsukan data penelitian, Memalsukan hasil, dan penipuan kepengarangan, yang dimana memiliki arti yaitu perampasan yang tidak semestinya atas konten dari milik penulis lain tanpa atribusi kredit yang semestinya. Selain itu, praktik-pratik terkutuk lainnya seperti redundansi dalam publikasi atau self-plagiarisme juga dipertimbangkan

dalam hal yang sama dalam kategori sebagai penanganan subjek penelitian yang ceroboh atau pembajakan (Krokosz, 2021, p. 3).

Jadi dapat diartikan plagiarisme adalah tindakan yang mengambil alih sebuah informasi, data, dan ilmu yang sebenarnya sebagai karya pribadi orang lain tanpa mengutip dari karya asli sebenarnya, Dengan banyaknya aksi plagiarisme tersebut, dikembangkanlah sebuah alat atau tools yang dapat mendeteksi aksi plagiarisme tersebut.

Tools plagiarisme ini dibuat untuk mendukung dunia pendidikan karena plagiarisme sendiri dapat merusak keadilan di dunia pendidikan. Dengan adanya tools plagiarisme ini dapat mengatasi aksi plagiarisme-plagiarisme yang banyak beredar. Aksi plagiarisme dapat dilihat dalam dunia akademik, plagiarisme sering terjadi dalam tugas mahasiswa yang dikumpulkan. Aksi-aksi seperti ini yang dapat merugikan dunia pendidikan.

Tools-tools yang dapat dipakai sering disebut sebagai aplikasi pendeteksi plagiarisme yang berupa *web*, aplikasi ini tersebar secara *online* di internet. Aplikasi pendeteksi plagiarisme telah digunakan oleh banyak kalangan dosen ataupun guru, contohnya seperti turnitin.

Oleh karena itu, Penelitian ini bertujuan untuk membuat sebuah aplikasi dengan bentuk *website* yang dapat mendeteksi plagiarisme. Aplikasi pendeteksi plagiarisme ini dibuat dengan menerapkan algoritma pencocokan *string* dalam dokumen teks untuk mencari kata-kata umum di antara dokumen, algoritma yang digunakan adalah *TF-IDF* (*term frequency-inverse document frequency*) dan *Cosine Similarity*. Dengan dilakukannya aksi pencocokan *string* dalam dokumen-dokumen yang dicocokkan, maka dihasilkan sebuah *output* yang dapat memberikan informasi seberapa dekat dokumen satu dengan dokumen lainnya. Berdasarkan latar belakang yang dijelaskan sebelumnya.

Maka diambilnya sebuah judul “Aplikasi Deteksi Plagiarisme Dokumen dengan Metode *TF-IDF* dan *Cosine Similarity* Berbasis *Web*”, semoga dengan aplikasi ini dapat membantu dosen dan mahasiswa untuk mengurangi tingkat atau aksi plagiarisme yang sering terjadi.

1.2 Identifikasi Masalah

1.2.1 Dalam dunia pendidikan masyarakat

Dalam dunia pendidikan ini kreatifitas sangat lah diperlukan, tetapi banyak terjadi kejadian yang mencuri hasil kreatifitas orang lain, ini sering disebut sebagai aksi plagiarisme. Aksi-aksi ini sangatlah memalukan dalam dunia pendidikan ini, jika hal ini terjadi terus-menerus dan tidak di hentikan maka, aksi seperti ini akan merugikan orang banyak.

1.2.2 Dalam dunia Ilmu Pengetahuan dan Teknologi (IPTEK)

Tidak hanya akan merugikan orang banyak, jika hal ini menyebar ke dunia ilmu pengetahuan dan teknologi (IPTEK), karya-karya original yang seharusnya dilakukan dengan penelitian-penelitian pribadi yang seharusnya dapat membantu perkembangan ilmu pengetahuan akan menghilang karena aksi-aksi plagiarisme. Hal ini dapat membuat masalah yang sangat besar, jika aksi plagiarisme ini terus berjalan, karya-karya original hasil penelitian pribadi tidak akan pernah ada lagi, karena takut akan terjadinya aksi plagiarisme terhadap karya-karya original tersebut.

1.3 Ruang Lingkup

Adapun batasan masalah yang didapat dalam penelitian ini, adalah: Metode yang dipakai terdiri dari 2 jenis, yaitu *TF-IDF* dan *Cosine Similarity*. Dua metode ini adalah inti dari penelitian ini, dua metode ini akan berkerja untuk menghitung persentase plagiarisme dalam satu dokumen. Data-data yang dipakai adalah data dari buku, jurnal atau artikel, dan *pdf-pdf* yang dikumpulkan dari *google*, yang dikumpulkan untuk

menjadi data uji, data-data ini akan dimasukkan ke *database* yang sudah disediakan.

Data uji yang akan digunakan berjumlah 20 data.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Adapun tujuan dibuatnya penelitian ini adalah sebagai berikut :

1. Untuk mengetahui hasil keefektifan dari metode algoritma *TF_IDF* dan *Cosine Similarity* dalam mendeteksi plagiarisme.
2. Agar dapat mengadaptasi aplikasi pendeteksi plagiarisme untuk digunakan dalam lingkungan akademik.
3. Agar dapat mengurangi aksi dari Plagiarisme.

1.4.2 Manfaat

Adapun manfaat dibuatnya penelitian ini adalah sebagai berikut :

1. Agar dapat meningkatkan efisiensi dalam menilai tugas mahasiswa sehingga dosen dan guru dapat memberikan umpan balik yang adil.
2. Untuk mendorong perkembangan dalam dunia pendidikan agar menjadi lebih baik dengan mengurangi aksi plagiarisme.
3. Berkurangnya aksi plagiat yang sangat banyak-banyak akhir-akhir ini.

1.5 Teknik Pengumpulan Data

1.5.1 Studi Pustaka

Dilakukan studi Pustaka dengan mengumpulkan jurnal, buku, dan artikel atau *pdf-pdf* dari berbagai sumber lainnya. Hal ini dilakukan untuk membuat dokumen-dokumen tersebut dapat dipakai untuk dijadikan sumber ide, dan juga dijadikan sebagai data penguji untuk menguji dokumen user dengan dokumen *admin* yang sudah di kumpulkan.

1.5.2 `Kuesioner

Dilakukannya kuesioner agar user dapat menilai antar muka dan proses yang dilakukan oleh aplikasi, dengan penilaian tersebut, maka akan diketahui apakah aplikasi yang akan dibuat nanti layak atau pun tidak layak menurut dari penilaian yang dihitung dari forum kuesioner yang dibuat.

1.6 Sistematisan Penelitian

Untuk lebih jelas dalam memahami Penelitian ini, Penelitian ini akan terbagi menjadi 5 bab dengan masing-masing uraian dengan garis besar yang dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini berisi pendahuluan mengenai latar belakang masalah, identifikasi masalah, ruang lingkup, tujuan dan manfaat, pengumpulan data.

BAB II LANDASAN TEORI

Pada bab ini berisi tentang teori-teori yang diambil dari jurnal-jurnal yang berkaitan dengan penyusunan penelitian serta beberapa literatur yang berhubungan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini dilakukannya proses analisa dan perhitungan manual yang akan dijadikan sebagai patokan untuk keakuratan aplikasi.

BAB IV PENGUJIAN DAN PENGIMPLEMENTASI

Pada bab ini akan memulai proses pembentukan aplikasi dari beberapa teori yang telah dipakai dan merancang antarmuka aplikasi yang diinginkan, dan menunjukkan aplikasi yang sudah dibentuk dari rancangan yang sudah dibuat.

BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi pembahasan kesimpulan dari penelitian serta saran yang diberikan terhadap penelitian yang telah dilakukan dan yang akan dilanjutkan jika ingin diangkat kembali sebagai penelitian.



BAB II

LANDASAN TEORI

2.1 Teori Umum

2.1.1 Data

Data adalah sebuah keterangan atau informasi yang didalamnya berbentuk angka bisa juga berbentuk kategori yang dihasilkan dari sebuah proses pengamatan, penghitungan, dan pengukuran suatu variabel yang menggambarkan sebuah masalah (Abdullah et al., 2022, p. 63).

Data sangat penting karena data menyediakan informasi yang diperlukan untuk mengidentifikasi, pola, dan hubungan antara variable-variabel yang relevan, dengan menganalisis sebuah data secara menyeluruh, kita dapat memperoleh informasi-informasi yang mungkin dibutuhkan untuk analisis selanjutnya.

Data-data yang didapatkan akan diolah dengan baik agar nantinya data tersebut dapat mudah diproses. Data juga dapat dikelompok menjadi 2 jenis data, yaitu :

1. Data primer

Data primer adalah sebuah data yang didapatkan melalui organisasi, lembaga, badan, institusi atau perorangan yang diambil langsung dari objeknya (Abdullah et al., 2022). Data ini biasanya didapatkan dari proses wawancara, data-data tersebut akan dikumpulkan menjadi satu dan akan membuat sebuah data primer yang akan digunakan untuk proses penelitian selanjutnya.

2. Data sekunder

Data sekunder adalah data yang dibutuhkan suatu organisasi atau perusahaan tetapi proses pengambilannya akan dilakukan kepada pihak lain yang dimana data tersebut sudah menjadi data jadi (Abdullah et al.,

2022), contoh dari data sekunder ini biasanya seperti data kependudukan, data jurnal, dan data tabel.

2.1.2 Informasi

Informasi adalah hasil yang didapatkan dari sebuah pemrosesan data yang relevan, data-data tersebut haruslah memiliki manfaat bagi yang memakai data tersebut (Lim & Ridho, 2021, p. 51).

Dengan demikian, informasi dapat dikatakan hasil pemrosesan data yang dapat memberikan nilai manfaat untuk yang memakai data tersebut. Informasi harus memenuhi beberapa persyaratan yang sesuai agar informasi tersebut bisa dijadikan sebagai informasi yang baik. Ada beberapa hal yang harus dipenuhi untuk menjadi informasi yang berkualitas (Lumbangaol & Ridho, 2020, p. 84), yaitu :

1. Akurat

Informasi tersebut haruslah benar dan tidak boleh adanya kesalahan. Harus secara cermat yang mengacu kepada tujuannya, dan sumber informasi tersebut haruslah sah agar informasi tersebut dapat disampaikan secara baik.

2. Tepat waktu

Informasi yang diberikan haruslah tepat waktu jangan sampai tertunda, karena informasi adalah aspek dasar dalam pengambilan sebuah keputusan, informasi-informasi yang tidak dilanjutkan maka tidak lagi memiliki arti atau berharga. keputusan selanjutnya maka bisa menjadi bencana bagi organisasi dan perusahaan.

3. Relevan

Informasi haruslah bermanfaat bagi yang memakainya. informasi-informasi tersebut akan menjadi penting bagi setiap orang berbeda.

Tujuan penerapan informasi ini adalah agar informasi-informasi yang diberikan kepada pengguna tersebut konkrit, tepat waktu, dan relevan, agar pengguna mendapatkan informasi yang pasti.

2.2 Teori Khusus

2.2.1 Plagiarisme

Plagiarisme adalah proses dilakukannya copy dan paste dari produk intelektual orang lain yang disalahgunakan tanpa menyebutkan nama penulis, penemu, dan penggagas orisinal (Shadiqi, 2019).

Plagiarisme dapat dijadikan 4 kategori, dan setiap kategori memiliki cara plagiarisme yang berbeda-beda. Dapat dilihat 4 kategori tersebut, yaitu:

a. *Verbatim Plagiarisme*

Dalam *Verbatim Plagiarisme* ini, penulis secara langsung mengambil karya atau pendapat orang lain tanpa mengubahnya, dapat dikatakan sama persis. *Verbatim Plagiarisme* sering dikenal sebagai *Copy-Paste*.

b. *Patchwork Plagiarisme*

Patchwork Plagiarisme adalah *Plagiarisme* berjenis penggabungan. Karena penulis akan melakukan penggabungan potongan ide atau karya orang lain yang akhirnya diakui menjadi gagasan atau ide sendiri. Pelaku pengguna *Patchwork Plagiarisme* ini sering menggabungkan potongan-potongan ide atau kalimat dari beberapa penulis yang akan dijadikan satu kalimat utuh tanpa mencatumkan sumber sama sekali.

c. *Paraphrase Plagiarisme*

Paraphrase Plagiarisme biasanya dilakukan dengan mengubah kalimat dari kata-kata aslinya, namun mengklaim bahwa kata-kata tersebut

sebagai tulisan atau ide sendiri. Tulisan dari kata-kata asli sering diubah sedemikian rupa agar tidak terkena plagiat. Pelaku Paraphrase Plagiarisme biasanya tidak jujur dan tidak mau mencantumkan sumber.

d. *Plagiarisme* kata atau Frase kunci

Pelaku pengguna *Plagiarisme* kata atau Frase kunci akan melakukan pengambilan kata atau frase kunci dari suatu tulisan. Pengambilan tulisan-tulisan tersebut akan disusun, kemudian dijadikan tulisan baru oleh pelaku.

Pelaku *plagiarisme* kata atau frase kunci ini akan menjabarkan dan merangkai tulisan-tulisan yang sudah diambil dalam suatu kalimat utuh untuk membentuk suatu makna baru tanpa menyebutkan sumber sama sekali.

2.2.2 *Text Mining*

Text Mining adalah proses memeriksa sejumlah besar dokumen untuk menemukan informasi-informasi baru atau membantu menjawab pertanyaan penelitian tertentu. mengidentifikasi fakta, hubungan, dan pernyataan yang tersisa. Setelah diekstraksi, informasi ini diubah menjadi bentuk terstruktur yang dapat dianalisis lebih lanjut atau disajikan secara langsung menggunakan tabel *HTML* yang dikelompokkan, peta pikiran, bagan, dll (Tojimamatov et al., 2023, p. 284)

Dalam Penelitian ini, *text mining* adalah inti dari semuanya karena dengan proses *text mining* semua data dapat diperoleh dan di *preprocessing*, walaupun *text mining* hanya 2 kata tetapi dunia *text mining* sangatlah luas. Banyak kegunaan yang dapat diperoleh dari *text mining*. *Text mining* terdiri dari 3 alat utama (Puri et al., 2023, pp. 5–6), yaitu:

1. *Extraction of Information*

Langkah ini dilakukan dengan cara mengikuti susunan teks yang telah ditentukan dengan mencocokkan polanya, Langkah ini adalah langkah awalan dimana seorang Penulis mengekstrak informasi yang relevan dari artikel Penelitian yang tidak terstruktur. Penulis biasanya memilih artikel yang sesuai, dan mencoba untuk menemukan serta mengumpulkan informasi dengan membedakan teks yang relevan.

2. *Discovery of Suitable Text*

Discovery of Suitable Text (DiscoText) adalah aspek paling penting dari *text mining*. Mengapa karena *DiscoText* sebuah kumpulan data terstruktur dari teks tidak terstruktur. Biasanya Penulis menggunakan alat *knowledge discovery in database (KDD)* atau penemuan pengetahuan dari database untuk memperoleh data terstruktur dan relevan yang dapat membantu dalam analisis selanjutnya. Penulis menggunakan ekstraksi kata kunci untuk mengklasifikasikan teks, membuat kelompok istilah, dan lain-lain.

3. *Text Analytics*

Text Analytics adalah proses otomatis yang membantu menafsirkan sejumlah besar teks tidak terstruktur menjadi data kualitatif, terutama untuk mengungkap wawasan, tren, dan pola. Langkah-langkah analisis teks dapat mengikuti langkah-langkah spesifik seperti berikut (Tojimamatov et al., 2023):

1. Analisis deskriptif adalah metode yang paling biasa atau bisa dikenal lebih umum, seperti "apa yang terjadi?", Penulis akan menjawab pertanyaan dan menganalisis data masuk dan historis secara real-

time. Tujuan utama dalam analisis deskriptif adalah untuk menentukan alasan dan pola keberhasilan atau kegagalan di bidang tertentu dan menggunakan informasi ini secara efektif.

2. Analisis prediktif akan melakukan proses kerja berdasarkan data yang tersedia. Analisis ini akan membantu memprediksi perkembangan peristiwa yang paling mungkin terjadi. Analisis prediktif menggunakan template siap pakai berdasarkan objek atau peristiwa apa pun yang memiliki kumpulan properti serupa.
3. Analisis preskriptif adalah tingkat yang lebih tinggi daripada analisis prediktif. Dengan bantuan Big Data dan teknologi modern, Penulis kemungkinan dapat mengidentifikasi titik masalah dalam bisnis atau aktivitas lainnya dan menghitung cara menghindarinya di masa depan.
4. Analisis diagnostik biasanya menggunakan data untuk menganalisis penyebab suatu insiden. Ini membantu untuk mengidentifikasi anomali dan hubungan acak antara peristiwa dan tindakan.

2.2.3 NLP (Natural language processing)

NLP adalah sebuah pemahaman bahasa alami yang membantu mesin "membaca" teks (atau masukan lain seperti ucapan) dengan menyimulasikan kemampuan manusia untuk memahami bahasa alami seperti Inggris, Spanyol, Mandarin, dan lain-lain. *NLP* mencakup pemahaman bahasa alami dan pembuatan bahasa alami, yang meniru kemampuan manusia untuk menghasilkan teks bahasa alami, seperti merangkum informasi atau terlibat dalam dialog.

Sebagai sebuah teknologi, *NLP* adalah teknologi yang terbaru, *NLP* telah dikembangkan selama bertahun-tahun, seperti *Siri*, *Alexa*, dan Pencarian *Google Voice*. Teknologi-teknologi tersebut menggunakan *NLP* untuk memahami dan merespons permintaan dari pengguna (Tojimatov et al., 2023).

2.2.4 TF-IDF

TF-IDF (*term frequency-inverse document frequency*) adalah sebuah perhitungan terhadap frekuensi kemunculan sebuah term pada dokumen yang ada. Algoritma dari *TF-IDF* akan melakukan pemeriksaan terhadap kemunculan tiap kata pada dokumen dari hasil *filtering*, tokenisasi. Rumus *TF-IDF* dapat dihitung sebagai, berikut:

$$W_{ij} = tf_{ij} * idf$$

Ket :

W_{ij} : Bobot kata *term* ke-j dan dokumen ke-i

tf_{ij} : Jumlah kemunculan kata / *term* ke-j dalam dokumen ke-i

idf : $\log N/df$

N : jumlah semua dokumen yang ada

n : jumlah dokumen yang mengandung term ke-j

2.2.5 Cosine Similarity

Cosine Similarity dalam Penelitian ini digunakan untuk proses perhitungan kesamaan dari dokumen. Adapun rumus dari *Cosine Similarity* adalah sebagai, berikut:

$$SV = \frac{AB}{\|A\|\|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Ket :

A : Vector A

B : Vector B

$\|A\|$: Panjang Vector A

$\|B\|$: Panjang Vector B

2.2.6 Waterfall

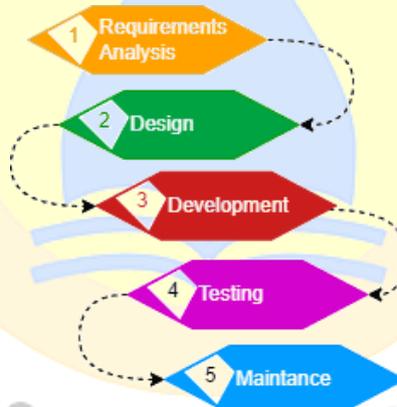
Waterfall adalah pendekatan yang terstruktur serta berurutan yang memungkinkan aplikasi tetap fokus pada setiap tahapan pengembangan dengan seksama, dari tahap awal yaitu perencanaan hingga tahap pengimplementasian (Madani & Haryono, 2023, p. 196). Dapat diartikan bahwa metode *waterfall* adalah sebuah metode pengembangan aplikasi yang didasari oleh urutan-urutan tertentu

2.3 Teori Analisa dan Perancangan

2.3.1 Teori Metodologi Penelitian

a. *Waterfall*

urutan-urutan dari *waterfall* memiliki beberapa langkah seperti diagram berikut ini (Ningki & Noviyanti, 2023, p. 109) :



Gambar 2.1 Metode Waterfall

Sumber : (Ningki dan Noviyanti, 2023, p. 109)

1. *Requirements Analysis*

Pada tahap pertama terjadi proses analisa aplikasi terhadap kebutuhan dari pengguna, proses ini akan dilakukan dengan cara menganalisa *website-website* yang memiliki sistem kerja yang sama seperti penelitian ini, analisa-analisa ini akan menjadi dokumen kebutuhan yang akan diperlukan untuk memaksimalkan kerja dari

aplikasi yang akan dibuat, seperti fitur dan fungsi yang akan dikembangkan pada sistem aplikasi.

2. *Design*

Pada tahap ini, terjadinya proses perancangan *design system* seperti perancangan antarmuka aplikasi, perancangan basis data, dan perancangan fungsi-fungsi yang akan diimplementasikan. Hasil dari tahap ini akan menghasilkan rancangan yang jelas untuk menciptakan arsitektur aplikasi yang detail.

3. *Development*

Setelah proses perancangan berhasil, dilakukannya proses pengimplementasian rancangan, proses-proses tersebut melibatkan *code-code* yang sudah ditentukan. Proses ini akan menghasilkan aplikasi yang berjalan seperti yang dirancang sedemikian rupa untuk menciptakan aplikasi plagiarisme berbasis *website*.

4. *Testing*

Setelah aplikasi sudah di *develop*, terjadi proses *testing* yang dimana proses pengujian akan dilakukan, proses ini bertujuan untuk memastikan sistem kerja aplikasi berfungsi dengan baik dan berjalan sesuai dengan persyaratan-persyaratan yang sudah ditentukan.

5. *Maintenance*

Setelah proses pengujian berhasil, maka akan dilakukannya proses *Maintenance*, hal ini dilakukan untuk memantau kinerja sistem, jika terjadi kesalahan atau *bug* yang sebelumnya tidak diketahui. Proses perbaikan akan dilakukan, hal ini dilakukan untuk memastikan *bug* tersebut tidak ada lagi untuk mengganggu dalam proses kerja

aplikasi. *Maintance* dilakukan juga untuk meningkatkan sistem kerja aplikasi agar menjadi lebih efisien.

b. Normalisasi Data

Normalisasi data adalah Metode *decimal scaling* yang merupakan metode transformasi data dengan normalisasi untuk menyamakan rentang nilai pada setiap atribut dengan skala tertentu dengan menggerakkan nilai desimal dari data ke arah yang diinginkan (Kusnaldi et al., 2022, p. 330).

c. *Preprocessing*

Preprocessing adalah tahap persiapan sebelum seluruh model dan algoritma selanjutnya diimplementasikan; misalnya segmentasi kata untuk bahasa Cina, Jepang, Vietnam, dan lainnya bahasa yang mungkin memerlukan segmentasi kata. (Zong et al., 2021, p. 11).

Preprocessing memiliki beberapa teknik utama (A. Hermawan et al., 2023, pp. 131–132), yaitu :

1. *Lowercasing*

Lowercasing dilakukan agar data-data yang akan digunakan untuk penelitian adalah huruf kecil tanpa ada menggunakan huruf besar, yang dapat dilihat pada Tabel 2.1 dibawah ini. Hal ini dilakukan karena komputer akan membedakan huruf jika teks huruf besar dan huruf kecil. Data latih yang akan dianalisis di penelitian ini akan memiliki huruf kecil semua, sehingga bila terdapat data dengan menggunakan huruf kapital maka akan terlewat atau tidak terdeteksi oleh data training yang ada untuk dipakai.

Tabel 2.1 *Lowercasing*

<i>Raw</i>	<i>Lower Cased</i>
AMAN	

AmAn
Aman

Aman

Sumber : (A. Hermawan et al., 2023)

2. Noise Removal

Noise removal ini dilakukan untuk menghilangkan simbol-simbol aneh yang ada sebelum dilakukan proses analisa lebih lanjut.

Contoh dari *Noise removal* dapat dilihat dari Tabel 2.2 berikut ini.

Tabel 2.2 *Noise Removal*

<i>Tweet</i>	<i>Tweet tanpa Noise</i>
@MVSOLAR Ada psbb segala sih ☐ 🤔 @steffislsbl Kalo di tempat gue udh ga psbb jd dah banyak yang nongki	mvsolar ada psbb segala sih steffislsbl alo di tempat gue udh ga psbb jd dah banyak yang nongki

Sumber : (A. Hermawan et al., 2023)

3. Stopword

Stopword mengandung kata-kata kurang berarti untuk dianalisa, hal ini dapat mengurangi dari kualitas data yang dimana akan dianalisa. Tujuan dari penghilangan *stopword* ini dilakukan untuk mengurangi jumlah dari teks yang akan diproses lebih lanjut nantinya, sehingga menyebabkan proses yang akan dilakukan nantinya menjadi lebih cepat dan tidak memakan banyak waktu. Untuk melakukan penghilangan atau menghilangkan *stopword* dalam penggunaan bahasa Indonesia maka menggunakan *library Python* Sastrawi.

4. Text Transformation

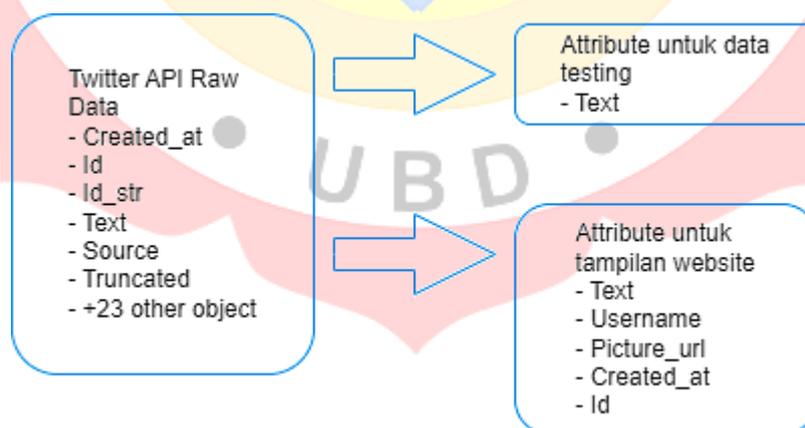
Text Transformation atau dikenal dengan Tokenisasi. Tokenisasi merupakan proses mengubah teks yang awalnya lebih besar menjadi sekumpulan beberapa teks yang kecil, ini disebut juga token. Token dapat berupa kata-kata, Kalimat, atau paragraf. Tokenisasi dapat

memisahkan antara tulisan yang ada dengan mencari atau menemukan pembatas dalam sebuah kata. Pembatasan kata yang umum atau biasa adalah tanda spasi atau *Whitespace*.

5. *Feature Selection*

Feature Selection adalah pemilihan atribut yang perlu sebelum dilakukan proses analisis dari data. Proses ini terjadi dikarenakan banyak data yang tersedia dan jumlahnya juga akan terus bertambah sehingga perlu untuk dilakukan *Feature Selection* agar dapat meningkatkan nilai dari akurasi, serta juga meningkatkan *performance* pada data yang berdimensi sangat tinggi.

Selain itu *Feature Selection* atribut dilakukan agar bentuk dari data-data yang akan dianalisa menjadi lebih sederhana dan juga mengurangi jumlah ukuran data yang akan diproses saat pengumpulan data. *Feature selection* dapat dilihat pada Gambar 2.2



Gambar 2.2 *Feature Selection*

Sumber : (A. Hermawan et al., 2023)

2.3.2 Teori Pemodelan Proses

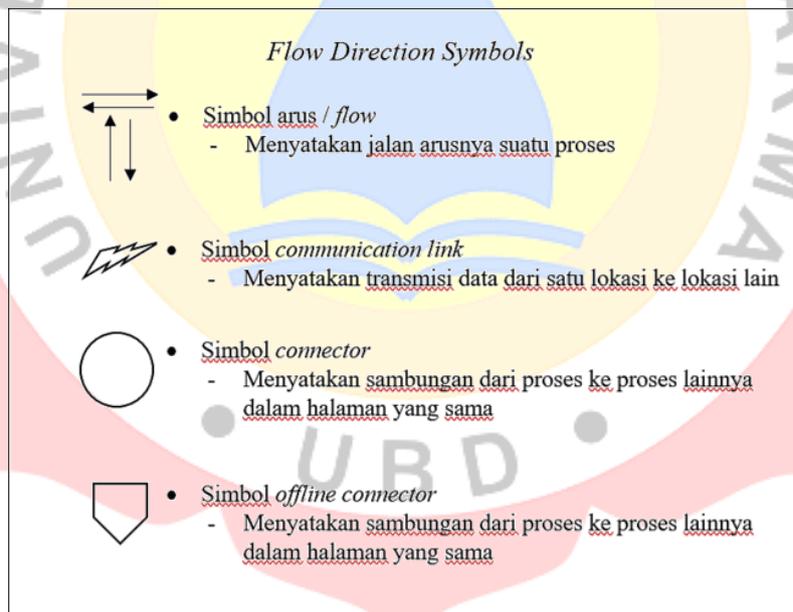
a. Flowchart

Flowchart adalah sebuah gambar yang berisi simbol-simbol, simbol-simbol ini akan menggambarkan sebuah urutan dari proses awal hingga proses akhir (Endra, 2020, p. 11).

Simbol-simbol *flowchart* sudah memiliki simbol standarnya tersendiri yang tidak dapat lagi diubah-ubah, simbol-simbol *flowchart* ini terdiri dari 3 jenis *flowchart* (Endra, 2020), yaitu :

1. Flow direction symbols

Flow direction symbols berfungsi sebagai penghubung simbol satu dengan simbol yang lain, simbol ini juga bisa disebut *connecting line*. Simbol-simbol tersebut dapat dilihat dari gambar berikut ini :



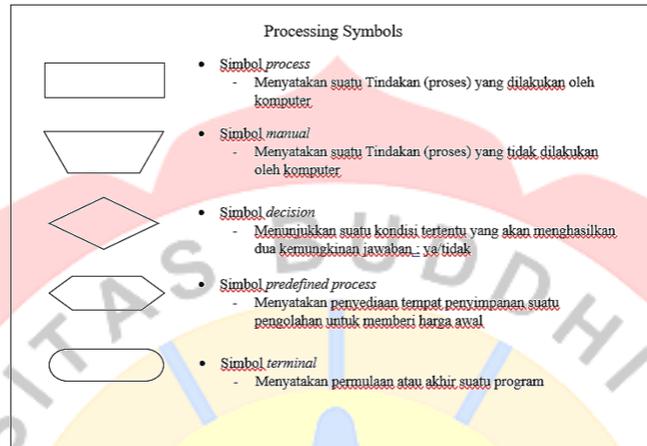
Gambar 2.3 *Flow Direction Symbols*

Sumber : (Endra, 2020)

2. Processing symbols

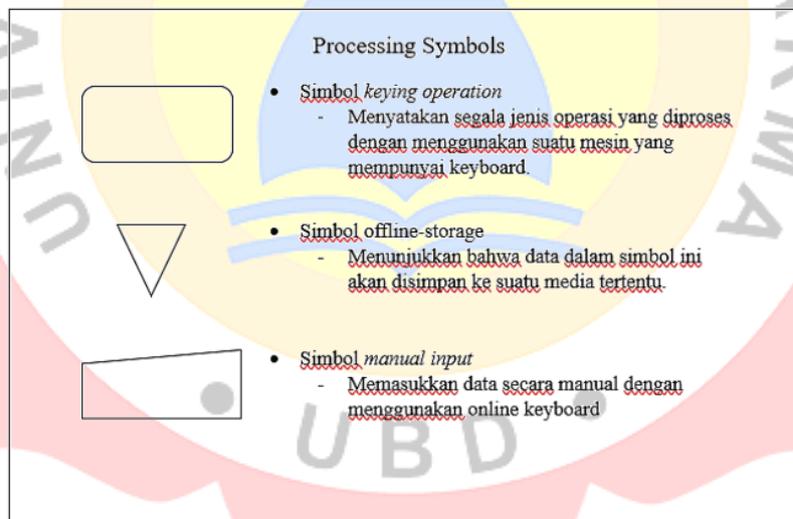
Simbol ini berfungsi untuk menunjukkan jenis-jenis operasi pengolahan seperti apa saja yang terjadi di dalam suatu proses/prosedur.

Simbol-simbol tersebut dapat dilihat dari gambar berikut ini :



Gambar 2.4 Processing symbols

Sumber : (Endra, 2020)

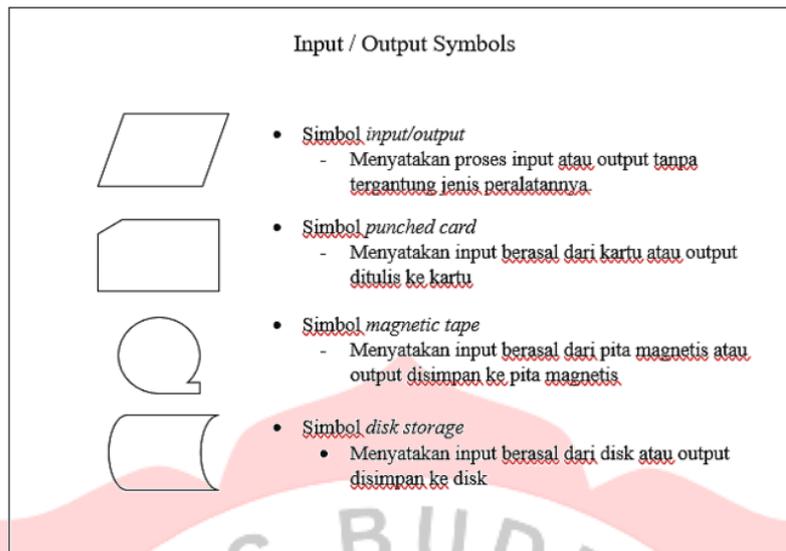


Gambar 2.5 Processing Symbols

Sumber : (Endra, 2020)

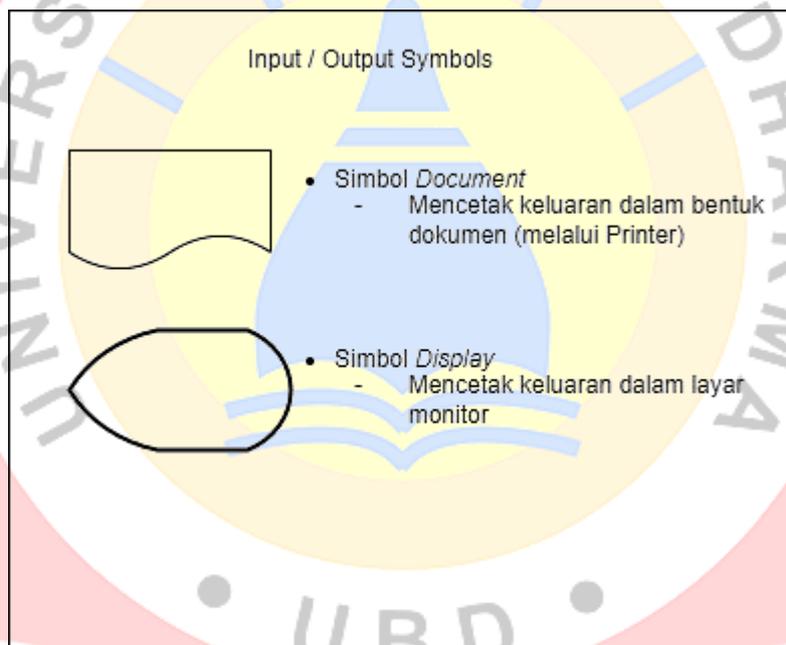
3. Input / Output symbols

Simbol ini akan menunjukkan jenis peralatan apa saja yang digunakan sebagai input atau output. Simbol-simbol tersebut dapat dilihat dari gambar berikut ini :



Gambar 2.6 *Input/Output Symbols*

Sumber : (Endra, 2020)



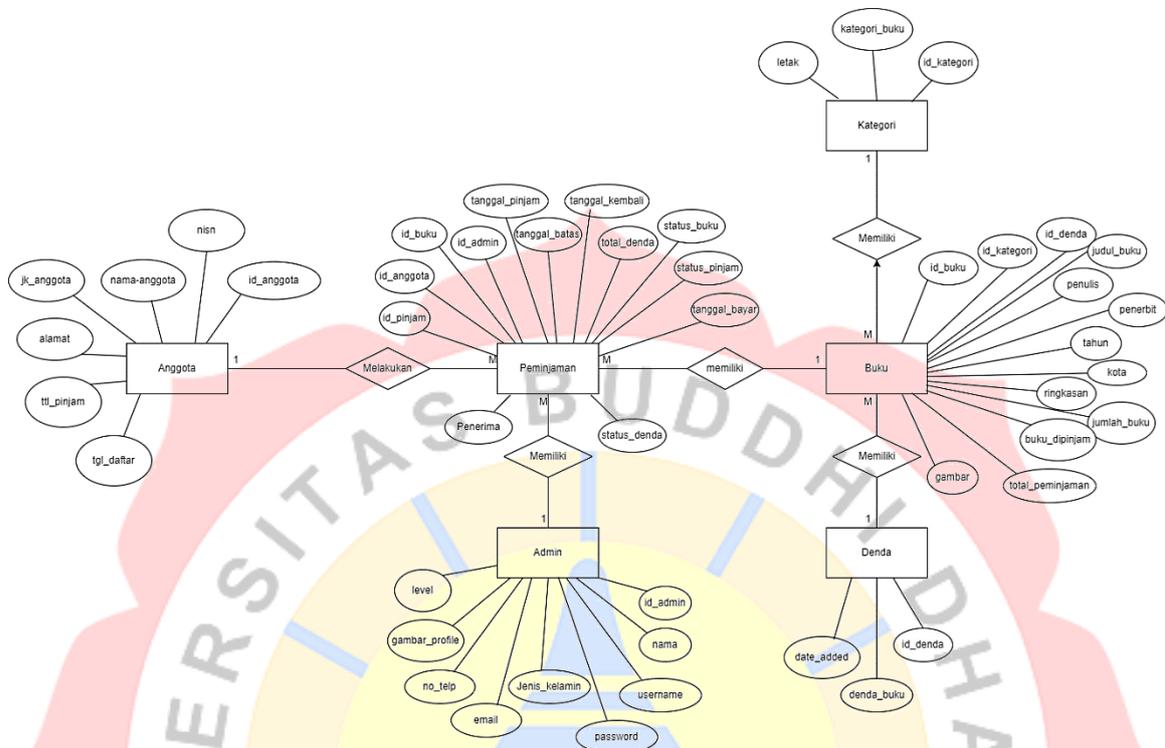
Gambar 2.7 *Input/Output Symbols*

Sumber : (Endra, 2020)

b. *ERD (Entity Relationship Diagram)*

ERD adalah suatu diagram yang berbentuk gambar atau simbol yang melakukan pengidentifikasian terhadap tipe dari entitas di dalam suatu sistem yang diuraikan didalam data dengan atribut-atributnya, serta menjelaskan

hubungan atau relasi diantara entitas tersebut (Hasanah & Untari, 2020, p. 37). Contoh *ERD* bisa dilihat pada gambar berikut ini:



Gambar 2.8 *Entity Relationship Diagram (ERD)*

Sumber : (Astuti & Samudra, 2023, p. 143)

c. *DFD (Data Flow Diagram)*

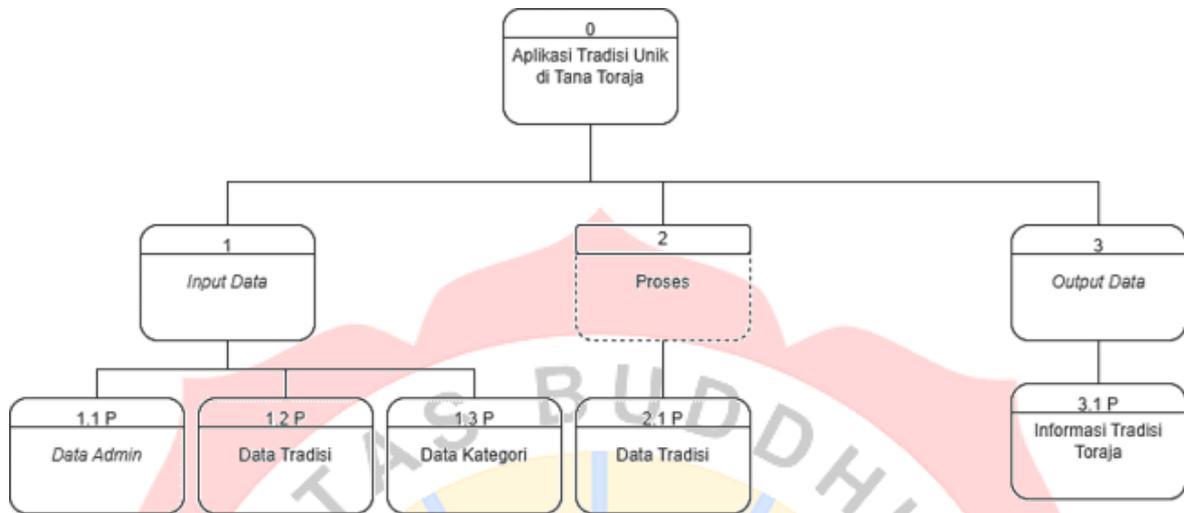
DFD adalah sebuah alat perancangan sistem yang berorientasi aliran data yang dimana memiliki konsep dekomposisi yang dapat digunakan untuk menjelaskan aspek analisis dan perancangan sistem yang mudah untuk dikomunikasikan kepada pengguna sistem (Wahyuni, 2023, p. 140). Dalam membuat perancangan *DFD* ada beberapa pedoman bagaimana menggambarkan *DFD* dengan baik , yaitu :

1. Bagan Berjenjang (*Hierarchy Chart*)

Bagan berjenjang (*Hierarchy chart*) biasanya digunakan untuk mempersiapkan penggambaran dari *DFD* ke level-level lebih bawah lagi. Bagan berjenjang dapat digambarkan dengan menggunakan

notasi proses yang biasa digunakan di *DFD* (Rozaq, 2020, p. 56).

Contoh dari bagan berjenjang dapat dilihat dari gambar 2.9.

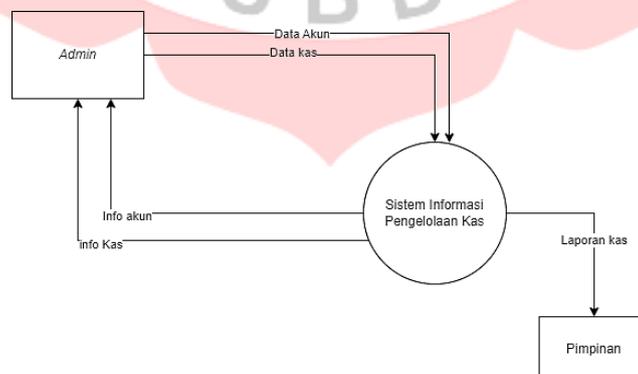


Gambar 2.9 Diagram Berjenjang

Sumber : (Oktavianus et al., 2023, p. 529)

2. Diagram Konteks (*Top Level*)

Diagram konteks adalah sebuah diagram yang terdiri dari proses-proses dan mewakili ruang lingkup dari sistem. Diagram konteks adalah level tertinggi dari *DFD*. Diagram konteks biasanya menggambarkan *inputan* atau *output* sistem di dalam sebuah sistem (Wahyuni, 2023). Contoh dari diagram konteks dapat dilihat dari gambar 2.10.

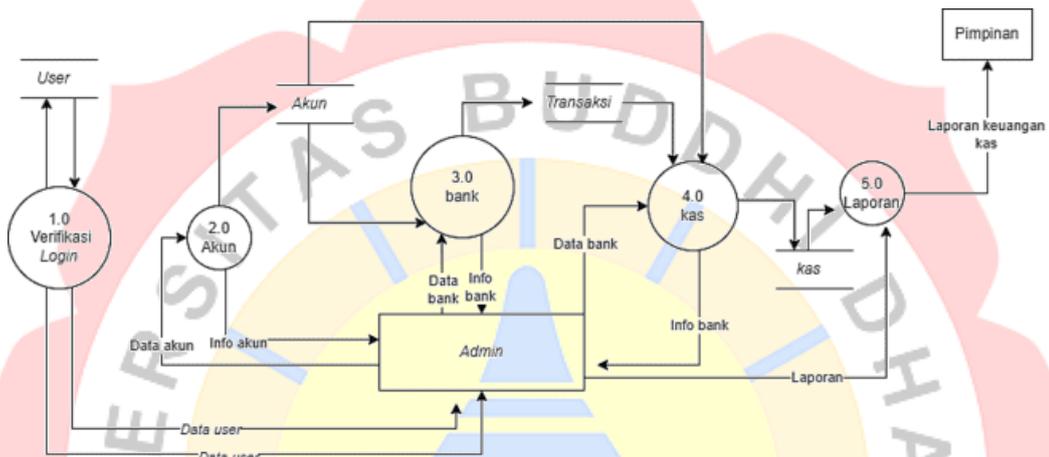


Gambar 2.10 Diagram Konteks

Sumber : (Wahyuni, 2023)

3. Diagram Overview (Level 0)

Diagram overview adalah gambaran yang memberikan pandangan secara singkat dan menyeluruh tentang sistem yang dicakup, gambaran yang menunjukkan fungsi atau proses utama yang terlibat dalam sistem (Wahyuni, 2023). Contoh dari *diagram overview* dapat dilihat dari gambar 2.11

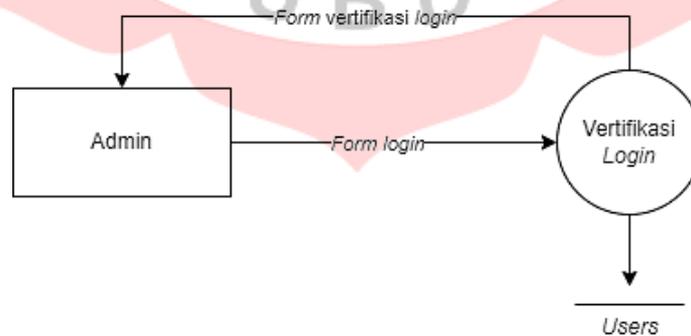


Gambar 2.11 Data Flow Diagram Level 0

Sumber : (Wahyuni, 2023)

4. Diagram Rinci (Level 1)

Diagram rinci adalah sebuah proses yang dipecah dari *level 0* menjadi proses yang lebih terinci lagi (Rozaq, 2020). Contoh dari diagram rinci dapat dilihat dari gambar 2.12

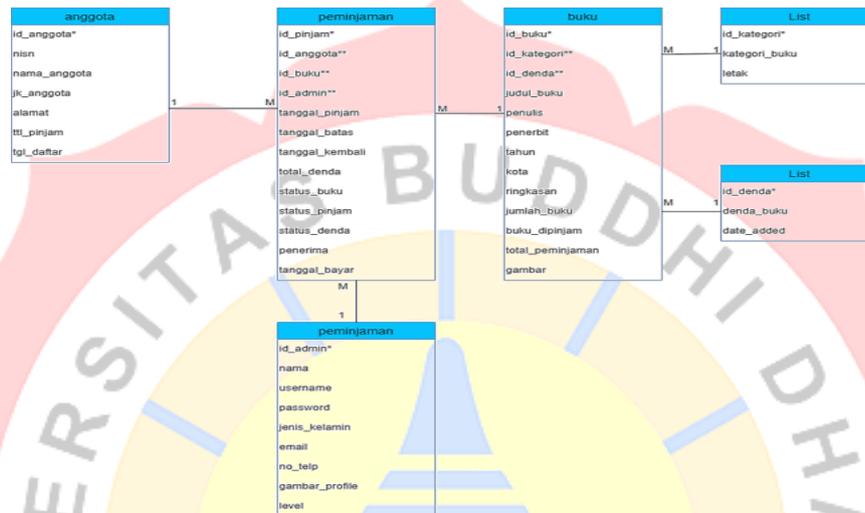


Gambar 2.12 Data Flow Diagram Level 1

Sumber : (Wahyuni, 2023)

d. *LRS (Logical record Structure)*

LRS adalah gambaran-gambaran dari hasil himpunan entitas atau hubungan yang terbentuk berdasarkan struktur-struktur *record* pada tabel-tabel di *database* (Astuti & Samudra, 2023). Contoh *LRS* bisa dilihat pada gambar berikut ini:



Gambar 2.13 *Logical record Structure (LRS)*

Sumber : (Astuti & Samudra, 2023)

2.3.3 Teori Aplikasi

1. *Visual Studio Code*

Visual studi code (Vs Code) adalah sumber aplikasi *open source* yang gratis dalam mengedit *code* pemrograman. *vs code* dikembangkan oleh *Microsoft*, aplikasi *vs code* adalah salah satu aplikasi *code editor* yang paling populer di lingkungan *development*, *vs code* memiliki banyak fitur yang dapat disesuaikan. *vs code* tidak hanya bagus untuk mengedit *code-code* tetapi juga mendukung kolaborasi dan lingkungan yang dihosting di *cloud* (Speight, 2021, p. 3).

Aplikasi ini dapat mendukung hampir semua Bahasa pemrograman seperti *php*, *javascript*, *python*, dan lain-lain, aplikasi *vs code* telah menjadi

salah satu aplikasi yang penting untuk para coder di dunia, karena aplikasi ini memiliki *UI (User Interface)* yang nyaman dan juga salah satu aplikasi yang dapat mengunduh *extension* yang diinginkan.

2. *Xampp*

Xampp adalah sebuah aplikasi yang dibentuk sedemikian rupa menjalankan peran sebagai *local web server*, ini berarti *xampp* akan memberikan sebuah *localhost* yang akan berperan sebagai *web server* dan juga sistem *database* (Remawati & Wijayanto, 2021, p. 25).

Xampp juga merupakan sebuah aplikasi yang didalam sudah tersedia beberapa kumpulan software seperti *PHPMyAdmin*, *Apache*, *MySQL*, *PHP*, *Perl*, dan lainnya, hal ini dapat memudahkan penggunaan karena tidak perlu lagi menginstall *software* satu persatu (Putri, 2022, p. 112).

3. *MySQL*

MySQL adalah sebuah *Database Management System (DBMS)* yang menggunakan perintah-perintah *Structured Query Language (SQL)* yang dimana sangat banyak digunakan saat ini dalam pembuatan aplikasi yang berbasis website. *MYSQL* dapat dibagi menjadi 2 lisensi, yaitu : pertama adalah *Free Software* yang dimana perangkat lunak dapat diakses oleh siapapun. Dan yang kedua adalah *shareware* dimana perangkat lunak yang berpemilik memiliki batasan dalam aspek penggunaannya (Remawati & Wijayanto, 2021).

4. *Python*

Sebuah aplikasi pasti membutuhkan program yang membuatnya berjalan, program-program tersebut dibentuk sedemikian rupa agar dapat

memproses keinginan dari pembuat, Dalam Penelitian ini akan menggunakan bahasa *Python*.

Python adalah bahasa yang cocok untuk belajar dan sebuah pemrograman yang nyata. *Python* sangat kuat dan sebuah bahasa pemrograman tingkat tinggi yang dibuat oleh seorang *programmer* bernama *Guido van Rossum* (Sharma et al., 2020, p. 115). Bahasa *Python* digunakan dalam penelitian ini dikarenakan dalam bahasa *python* terdapat modul-modul *library* yang dibutuhkan untuk meningkatkan proses perhitungan dalam pemrograman.

5. *Javascript*

Javascript adalah bahasa pemrograman, sebagian besar digunakan untuk mendefinisikan perilaku interaktif halaman *web*. *JavaScript* memungkinkan Anda membuat halaman web lebih interaktif dengan mengakses dan memodifikasi konten dan gaya halaman web saat sedang dilihat di *browser* (Dorman, 2020, p. 59). *Javascript* dapat digunakan Ketika menambahkan tag `<script></script>` didalam *tag* tersebut akan dijalankan proses pemrograman *javascript*-nya, seperti penjelasan *Javascript*, *web* akan terlihat indah jika pengguna dapat melakukan banyak interaksi dengan *web* kita, agar *web* tersebut tidak membosankan dan memiliki tampilan yang menarik, maka *Javascript* perlu digunakan.

6. *HTML (Hypertext Markup Language)*

HTML adalah format data yang digunakan untuk menyandikan konten dan struktur halaman *web* (Dorman, 2020). *HTML* ini memiliki ekstensi file seperti Namanya, yaitu *.html*. *code-code* dalam *HTML* biasanya ditransfer dan diterjemahkan ke dalam gambar secara visual, seperti *web* yang kita

sering lihat, tetapi biasanya html akan digabungkan dengan *code-code* bahasa pemrograman *javascript* dan *code-code css* untuk memperindah tampilan *web*.

HTML ini dapat dikatakan adalah hal dasar dalam pembuatan sebuah *web*, kerangka dasar dalam *website* adalah *HTML* ini. Jika *HTML* tidak ada maka tampilan akan berupa tampilan kosong berwarna putih tanpa ada kata-kata atau gambar.

7. *CSS (Cascading Style Sheet)*

CSS adalah sebuah penyedia gaya dan sesuai dengan namanya *css* adalah singkatan dari *cascading style sheet*, dan ini menggambarkan bagaimana halaman *web* seharusnya terlihat: ia mengatur warna, font, spasi, dan banyak lagi. Singkatnya, Anda dapat membuat situs web Anda terlihat sesuai keinginan (Casabona, 2020).

Sesuai pengertian Casabona, *CSS* dapat diartikan sebagai bahasa pemrograman khusus yang menyediakan segala hal tentang *design* sebuah *website*, bahasa pemrograman ini dapat membantu pembuat *website* untuk mendesain *websitenya* agar *UI (User Interface)* dari *website* tersebut nyaman untuk dilihat, sama seperti *javascript* untuk membuat tampilan *website* menarik, tapi *system* kerjanya berbeda.

Proses kerja *CSS* hanya membuat *design website* menarik, dari animasi, tampilan warna, dan ukuran dari gambar atau huruf. Proses pemanggilan dari *CSS* dapat berupa berikut :

```
<link rel="stylesheet" href="nama file.css" >
```

Sebenarnya, banyak proses pemanggilan dari *CSS* tetapi proses inilah yang paling rapi, Mengapa proses ini paling rapi karena proses pengerjaan pemrogramannya di berbeda *file*. Dengan berbeda *file* pembuat sebuah

website dapat memfokuskan proses pengerjaan design di hanya satu *file* saja, tetapi balik lagi semua sesuai kemauan sendiri.

2.4 Teori pengujian

2.4.1 Blackbox Testing

Blackbox testing adalah adalah sebuah pengujian yang dibentuk sebagai hasil dari eksekusi sebuah aplikasi melalui data-data yang diuji untuk memastikan fungsional dari aplikasi yang dijalankan, proses ini dilakukan untuk memverifikasi apakah aplikasi tersebut sudah sesuai dengan persyaratan atau belum (Huda et al., 2022, p. 122). Dalam *blackbox* ini terdapat beberapa metode pengujian, seperti *Sample Testing*, *Boundary Value Analysis*, dan *Equivalence Partitions*.

Dalam penelitian ini metode pengujian *blackbox* yang akan digunakan adalah *Equivalence partitions*. *Equivalence partitions* adalah proses sebuah pengujian yang berdasarkan inputan data pada setiap *form* yang dibuat pada aplikasi, setiap menu inputan akan dilakukan pengujian dan akan dikelompokkan dari fungsinya baik itu bernilai valid ataupun tidak valid (Suyudi et al., 2020, p. 199). Contoh teknik *black box Equivalence partitions* bisa dilihat di gambar berikut ini.

Tabel 2.3 Tabel Uji Black Box

No.	Form/Tampilan Yang Di uji	Jumlah Butir Uji	Hasil		Jumlah Bukti
			Sesuai	Tidak Sesuai	
1	<i>Login</i>	2	2	0	2
2	<i>Sub menu master buku</i>	7	7	0	7
3	<i>Submenu Data Master Anggota</i>	7	7	0	7
4	<i>Submenu Data Master Kategori</i>	7	7	0	7
5	<i>Submenu Data Master Penerbit</i>	7	7	0	7
6	<i>Subemenu Data Master Rak</i>	6	5	1	6
7	Menu Pengadaan Buku	5	5	0	5
8	<i>Submenu Data Transaksi Peminjaman Buku</i>	7	6	1	7

9	Submenu Data Transaksi Pengembalian Buku	6	5	1	6
10	Menu Data Pengguna	6	5	1	6
11	Menu Laporan Pengadaan Buku	4	3	1	4
12	Menu Laporan Peminjaman Buku	4	3	1	4

Sumber : (Azhari et al., 2023, pp. 378–379)

2.5 Tinjauan Studi

- a. (Fadhullah et al., 2022), Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi).

Tabel 2.4 Jurnal ke-1

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Aplikasi Deteksi Dini Plagiarisme Penelitian Ilmiah Menggunakan Algoritma <i>Cosine Similarity</i> Berbasis <i>Web</i>
2	Jurnal	Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) E-ISSN : 2580-1643
3	Volume, Nomor, dan Halaman	Volume 6, Nomor 3, Halaman 325-334
4	Bulan dan Tahun	Juli 2022
5	Penulis	Ahlun Nazar Fadhullah, Fauziah, Winarsih
6	Penerbit	<i>Research Division, KITA Institute</i>
7	Tujuan Penelitian	a. Meminimalisir terjadinya plagiarisme b. Memberikan solusi dalam menangani masalah plagiarisme dengan metode <i>Cosine Similarity</i>
8	Lokasi dan Objek Penelitian	Lokasi di Universitas dan objeknya adalah Pengembangan aplikasi berbasis <i>web</i> untuk mendeteksi dini plagiarisme dalam penelitian ilmiah menggunakan metode <i>Cosine Similarity</i>
9	Metode yang digunakan	a. Metode <i>Cosine Similarity</i> b. Metode <i>Preprocessing</i> seperti: <i>Noise Text Remover, Case Folding, Stemming, Stopword</i> .
10	Perancangan Sistem	a. Menggunakan <i>Use Case Diagram</i> untuk membuat aplikasi b. Menggunakan <i>flowchart</i> dalam membuat alur aplikasi c. Menggunakan teknik <i>preprocessing</i> di aplikasi

11	Hasil Penelitian	Hasil dari penelitian ini adalah : a. Tampilan awal berupa menu <i>login</i> b. Tampilan menu utama berupa menu yang memiliki beberapa fitur yang ada dalam program c. Tampilan menu pengguna d. Tampilan menu data latih e. Tampilan Menu Pengujian f. Tampilan Menu Klasifikasi g. Program dapat berjalan dengan baik dengan mengeluarkan hasil 31.183724712497%.
12	Kekuatan Penelitian	Kekuatan penelitian ini adalah : a. Penelitian ini menunjukkan <i>flowchart</i> dari program b. Penelitian ini menunjukkan <i>Use Case Diagram</i> c. Penelitian ini menggunakan algoritma <i>Cosine Similarity</i> yang dapat dikatakan berhasil. d. Penelitian ini menampilkan proses dari menu login hingga menu klasifikasi e. Penelitian ini menampilkan proses klasifikasi yang terjadi.
13	Kelemahan Penelitian	Kelemahan penelitian ini adalah : a. Tidak memberi tau bahasa pemrograman yang digunakan b. Tidak menjelaskan <i>database</i> yang digunakan c. Tidak menjelaskan secara rinci metode apa saja yang digunakan.
14	Kesimpulan	Kesimpulan : Aplikasi dapat dikatakan berjalan dengan baik menggunakan metode <i>Cosine Similarity</i> yang mendapatkan hasil 31.183724712497%, tetapi dalam penjelasan yang diutarakan dalam penelitian ini masih kurang rinci dan jelas.

b. (Herianto et al., 2021), Jurnal Ilmu Komputer (Computer Science Journal).

Tabel 2.5 Jurnal ke-2

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Aplikasi Deteksi Plagiarisme Judul Tugas Akhir Berbasis <i>Web</i> Dengan Menggunakan Algoritma <i>Rabin-Karp</i>

		<i>Rolling Hash</i> (Studi Kasus: Amik Mahaputra Riau)
2	Jurnal	Jurnal Ilmu Komputer (<i>Computer Science Journal</i>) E-ISSN : 2579 - 3918 P-ISSN : 2302 - 710X
3	Volume, Nomor, dan Halaman	Volume 10, Nomor 2, Halaman 107-112
4	Bulan dan Tahun	Oktober 2021
5	Penulis	Herianto, Yulisman, Winda Herrianti Manullang, Yuda Irawan
6	Penerbit	Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Hang Tuah Pekanbaru
7	Tujuan Penelitian	<ul style="list-style-type: none"> a. Mengembangkan aplikasi deteksi plagiarisme berbasis <i>web</i> menggunakan algoritma <i>Rabin-Karp</i> b. Untuk membantu administrator, kepala departemen, dan mahasiswa di perguruan tinggi riau untuk mendeteksi plagiarisme pada tugas akhir
8	Lokasi dan Objek Penelitian	Lokasi penelitian ini di Perguruan tinggi Riau dan Objek penelitiannya adalah Pengembangan aplikasi deteksi plagiarisme berbasis <i>web</i> dengan menggunakan algoritma <i>Rabin-Karp Rolling Hash</i> .
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Metode <i>Waterfall</i> b. Metode Algoritma <i>Rabin-Karp Rolling Hash</i> c. Metode <i>Unified Modeling Language (UML)</i>
10	Perancangan Sistem	<ul style="list-style-type: none"> a. <i>Unified modeling language (UML)</i> b. Menggunakan <i>activity diagram</i> untuk menggambarkan aliran kerja aplikasi.
11	Hasil Penelitian	<p>Hasil penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Tampilan menu <i>login</i> b. Tampilan Halaman Awal Pengguna c. Tampilan <i>Input</i> Mahasiswa d. Tampilan Daftar Data Mahasiswa e. Tampilan <i>Menu Input</i> Data Dosen f. Tampilan <i>Menu Input</i> Ketua Jurusan g. Tampilan Data Pengajuan Tugas Akhir h. Tampilan Laporan Cetak Plagiarisme i. Tampilan Daftar Data Tugas Akhir

		<ul style="list-style-type: none"> j. Tampilan Menu Cek Plagiarisme Tugas Akhir k. Tampilan Buk Terdeteksi Plagiarisme l. Tampilan Menu Laporan Pembayaran
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Menggunakan metode Algoritma <i>Rabin-Karp Rolling Hash</i> b. Perancangan system menggunakan metode <i>Unified Modeling Language (UML)</i> c. Mahasiswa dapat mengupload tugas akhir, menunggu proses, dan melihat hasil persentase plagiarisme. d. Ketua jurusan dapat mencetak hasil plagiarisme dan melakukan bimbingan terhadap tugas akhir mahasiswa.
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Tidak dijelaskan menggunakan Bahasa pemrograman apa b. Tidak menjelaskan dengan baik evaluasi kinerja program, karena tidak menjelaskan dengan baik hasil dari percobaan evaluasi kerja. c. Tidak menjelaskan dengan baik <i>database</i> yang digunakan
14	Kesimpulan	<p>Kesimpulan :</p> <p>Aplikasi dikatakan berkerja dengan baik dan memiliki <i>UI</i> yang menarik dan lengkap tetapi dalam studi kasus aplikasi ini hanya dapat digunakan di sekitaran kampus Amik Mahaputra Riau jika ingin disebar secara luas, maka akan dilakukan perombakan secara menyeluruh dengan <i>User Interface</i> dari aplikasi tersebut</p>

c. (Zulfahmi & Fitriyani, 2023), Jurnal Real Riset.

Tabel 2.6 Jurnal ke-3

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Aplikasi Deteksi Plagiarisme Skripsi Mahasiswa Fakultas Teknik Universitas Jabal Ghafur Menggunakan Algoritma <i>Winnowing</i> Dan <i>Jaccard Similarity</i> Berbasis Web
2	Jurnal	Jurnal Real Riset

		ISSN : 2685-1024 eISSN : 2774-7263
3	Volume, Nomor, dan Halaman	Volume 5, Nomor 1, Halaman 44-48
4	Bulan dan Tahun	Januari 2023
5	Penulis	Zulfahmi, Fitriyani
6	Penerbit	LPPM Universitas Jabal Ghafur
7	Tujuan Penelitian	<ul style="list-style-type: none"> a. Untuk mengembangkan sebuah aplikasi pendeteksi plagiarisme skripsi mahasiswa di Fakultas Teknik Universitas Jabal Ghafur b. Untuk memudahkan dalam mengecek tingkat persentase kesamaan dokumen skripsi sehingga plagiarisme dapat dihindari. c. Untuk membantu pembimbing dalam menyetujui skripsi yang ditulis oleh mahasiswa.
8	Lokasi dan Objek Penelitian	Lokasi penelitian ini adalah Fakultas Teknik Universitas Jabal Ghafur dan objek penelitiannya adalah pembuatan aplikasi plagiarisme <i>detection</i> dengan data skripsi-skripsi sebelumnya yang ada di perpustakaan Fakultas Teknik Universitas Jabal Ghafur.
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Studi Literatur b. Analisis Sistem dengan Penelitian ke Lapangan c. Metode Algoritma <i>Winnowing</i> dan <i>Jaccard Similarity</i> d. Metode Pengujian dan Analisis sistem
10	Perancangan Sistem	<ul style="list-style-type: none"> a. Bahasa pemrograman <i>HTML</i>, <i>CSS</i>, <i>PHP</i>, <i>Javascript</i> dan <i>Bootstrap</i> b. Menggunakan <i>database sql</i>
11	Hasil Penelitian	<p>Hasil penelitian ini adalah :</p> <ul style="list-style-type: none"> a. <i>Entity Relationship Diagram</i> (ERD) b. Rancangan halaman <i>login admin</i> c. Rancangan halaman data tahun d. Rancangan halaman data akun mahasiswa e. Rancangan halaman data penelitian sebelumnya f. Rancangan halaman <i>login</i> mahasiswa g. Rancangan halaman proses deteksi plagiat skripsi h. Rancangan informasi data penelitian sebelumnya i. Rancangan informasi hasil deteksi plagiat skripsi j. Tampilan halaman <i>login admin</i>

		<ul style="list-style-type: none"> k. Tampilan halaman data tahun l. Tampilan halaman data akun mahasiswa m. Tampilan halaman data penelitian sebelumnya n. Tampilan halaman <i>login</i> mahasiswa o. Tampilan halaman proses deteksi plagiat skripsi p. Tampilan informasi data penelitian sebelumnya q. Tampilan informasi hasil deteksi plagiat skripsi
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Mempermudah pembimbing dalam menyetujui skripsi yang ditulis oleh mahasiswa. b. Dilengkapi dengan halaman <i>login admin</i> yang berfungsi sebagai halaman keamanan khusus untuk pembimbing skripsi memasukkan dokumen skripsi ke dalam formulir dan menampilkan hasil deteksi plagiarisme. c. Kombinasi teknologi <i>web, Bootstrap</i>, dan <i>database MySQL</i> dapat mendukung fungsionalitas dan tampilan aplikasi dengan baik.
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Tidak menjelaskan secara rinci perancangan <i>database</i> b. Tidak menjelaskan secara rinci dari algoritma <i>Winnowing</i> dan <i>Jaccard Similarity</i> c. <i>Database</i> yang digunakan adalah <i>sql</i> tapi tidak ada penjelasan lebih lanjut tentang <i>database sql</i> d. Hasil dari perhitungan tidak dijelaskan secara rinci
14	Kesimpulan	<p>Kesimpulan :</p> <p>Aplikasi ini cukup berguna dalam melakukan pendeteksian plagiarisme tetapi dalam penelitian ini tidak dijelaskan secara rinci hasil dari plagiat yang didapatkan, maka tidak diketahui berapa persen yang dikeluarkan oleh perhitungan algoritma <i>Winnowing</i> dan <i>Jaccard Similarity</i> dalam penelitian ini.</p>

d. (L. Hermawan & Ismiati, 2020), Jurnal Buana Informatika.

Tabel 2.7 Jurnal ke-4

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Aplikasi Pengecekan Dokumen Digital Tugas Mahasiswa Berbasis <i>Website</i> .
2	Jurnal	Jurnal Buana Informatika ISSN : 2087-2534 eISSN : 2089-7642
3	Volume, Nomor, dan Halaman	Volume 11, Nomor 2, Halaman 94-103
4	Bulan dan Tahun	Oktober 2020
5	Penulis	Latus Hermawan, Maria Bellanier Ismiati
6	Penerbit	Universitas Atma Jaya Yogyakarta
7	Tujuan Penelitian	a. Untuk memberikan informasi tentang sejauh mana dokumen tugas mahasiswa memiliki kesamaan dengan dokumen tugas kuliah yang lain. b. Untuk membantu dosen dan mahasiswa dalam mengurangi tingkat plagiarisme.
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	a. Studi Pustaka b. Metode <i>TF-IDF</i> dan <i>Cosine Similarity</i> c. Metode <i>waterfall</i> d. Metode <i>text preprocessing</i>
10	Perancangan Sistem	a. Bahasa pemrograman <i>HTML</i> , <i>CSS</i> , <i>PHP</i> , dan <i>Javascript</i> b. Menggunakan database <i>MySQL</i> c. Menggunakan aplikasi <i>sublime</i> dan <i>xampp</i>
11	Hasil Penelitian	Hasil penelitian ini adalah : a. Tampilan <i>login</i> simpel <i>plagiarisme checker application</i> b. Halaman <i>upload</i> tugas mahasiswa c. Tampilan halaman hasil plagiarisme (Terdeteksi) d. Tampilan halaman hasil plagiarisme (tidak terdeteksi)
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a. Menggunakan algoritma pencocokan string seperti <i>TF-IDF</i> dan <i>Cosine Similarity Index</i> untuk menganalisis dan membandingkan dokumen teks dalam mencari kata-kata umum di antara dokumen tugas kuliah. b. Dapat menampilkan persentase plagiarisme beserta statusnya apakah terdeteksi plagiat atau tidak.

		c. Penggunaan dataset tugas mahasiswa yang cukup banyak (79 file) untuk melakukan analisis dan pengujian aplikasi deteksi plagiarisme.
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a. Tidak dijelaskan secara rinci metode-metode yang digunakan. b. Perancangan <i>database</i> yang tidak terlalu dijelaskan. c. Menggunakan <i>database SQL</i> tapi tidak terlalu dijelaskan.
14	Kesimpulan	Kesimpulan : Aplikasi dapat dikatakan berjalan dengan baik, tetapi penjelasan dari penelitian ini masihlah kurang rinci. Penjelasan-penjelasan hanya merujuk kepada <i>TF-IDF</i> dan <i>Cosine Similarity</i> sedangkan penjelasan dari <i>text preprocessing</i> masihlah kurang jelas.

e. (Nawawi et al., 2019), Jurnal Nasional Pendidikan Teknik Informatika.

Tabel 2.8 Jurnal ke-5

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Deteksi Plagiarisme Pada Dokumen Skripsi Berdasarkan Tingkat Kesamaan Dengan Menggunakan Metode <i>Longest Common Subsequence</i>
2	Jurnal	Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI) ISSN : 2089-8673 eISSN : 2584-4265
3	Volume, Nomor, dan Halaman	Volume 8, Nomor 3, Halaman 217-226
4	Bulan dan Tahun	Desember 2019
5	Penulis	Imam Nawawi, Putra Prima Arhandi, Faisal Rahutomo
6	Penerbit	Universitas Pendidikan Ganesha
7	Tujuan Penelitian	a. Untuk membuat aplikasi deteksi plagiarisme pada dokumen skripsi dengan menggunakan metode <i>Longest Common Subsequence</i> (LCS). b. Untuk membantu mahasiswa dan panitia Laporan Akhir dan Skripsi dalam mengukur tingkat plagiarisme pada dokumen tersebut.

8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Studi pustaka b. Metode <i>Longest Common Subsequences (LCS)</i> c. Metode pengolahan data, seperti <i>Convert</i> dokumen menjadi teks, <i>Filter</i> kandidat pembanding d. Metode <i>Model View Controller (MVC)</i>
10	Perancangan Sistem	<ul style="list-style-type: none"> a. Menggunakan metode <i>Model View Controller (MVC)</i> dalam pembuatan aplikasi b. Aplikasi akan mengubah data <i>pdf/doc</i> ke bentuk teks
11	Hasil Penelitian	<p>Hasil dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Kasus pertama b. Kasus kedua c. Kasus ketiga <p>Ketiga kasus tersebut memberikan jawaban yang sama sebesar 100%</p>
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah</p> <ul style="list-style-type: none"> a. Pengembangan aplikasi deteksi plagiarisme dokumen skripsi menggunakan metode <i>Longest Common Subsequence (LCS)</i>. b. Memungkinkan pengujian antara dua atau lebih dokumen pembanding c. Dapat mendeteksi plagiarisme dalam lebih dari satu kalimat dan dengan lebih dari satu kandidat pembanding.
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah</p> <ul style="list-style-type: none"> a. Tidak menjelaskan secara detail perancangan aplikasi b. Tidak menjelaskan Bahasa pemrograman yang dipakek c. Pemakaian <i>database</i> tidak dijelaskan secara detail
14	Kesimpulan	<p>Kesimpulan :</p> <p>Metode yang digunakan lumayan menarik tetapi penjelasan dan perancangan aplikasi masih kurang jelas, walaupun begitu hasil yang didapatkan dikategorikan akurat, tetapi dalam penjelasannya proses <i>input</i> sampai output masih membingungkan.</p>

f. (Ardhiansyah & Pradana, 2023), OKTAL: Jurnal Ilmu Komputer dan Science.

Tabel 2.9 Jurnal ke-6

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Deteksi Plagiarisme Pada <i>File</i> Dokumen Berdasarkan Tingkat Kesamaan Dengan Menggunakan Metode Algoritma <i>Rabin-Karp</i> Berbasis <i>Web</i>
2	Jurnal	OKTAL: Jurnal Ilmu Komputer dan Science eISSN : 2828-2442
3	Volume, Nomor, dan Halaman	Volume 2, Nomor 3, Halaman 760-766
4	Bulan dan Tahun	Maret 2023
5	Penulis	Hardison Pradana, Maulana Ardhiansyah
6	Penerbit	CV. Multi Kreasi Media
7	Tujuan Penelitian	a. Untuk merancang dan membangun sebuah aplikasi yang menggunakan algoritma <i>Rabin-Karp</i> untuk mendeteksi plagiarisme pada dokumen teks. b. Untuk menemukan persentase kemiripan antara dokumen asli dan dokumen yang diuji dengan menggunakan algoritma <i>Rabin-Karp</i> .
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	a. Studi pustaka b. Studi literatur c. Metode <i>waterfall</i> d. Metode algoritma <i>Rabin-Karp</i>
10	Perancangan Sistem	a. Menggunakan <i>database</i> b. Menggunakan metode <i>waterfall</i> dalam membuat aplikasi.
11	Hasil Penelitian	Hasil penelitian ini adalah : a. Tampilan halaman <i>login</i> b. Tampilan halaman <i>dashboard</i> c. Tampilan halaman kelola dokumen d. Tampilan halaman analisa dokumen e. Tampilan halaman hasil Analisa f. Tampilan halaman hasil analisa lanjut
12	Kekuatan Penelitian	Kekuatan dari Penelitian ini adalah : a. Penelitian ini memberikan penjelasan tentang proses <i>text preprocessing</i> b. Penelitian ini mengimplementasikan algoritma <i>Rabin-Karp</i> melalui metode pencocokan string untuk menghitung persentase kemiripan antara dua file dokumen teks.
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah

		<ul style="list-style-type: none"> a. Tidak memberikan informasi tentang bahasa pemrograman yang digunakan dalam pengembangan aplikasi. b. Tidak menyebutkan jenis <i>database</i> yang digunakan dalam sistem.
14	Kesimpulan	<p>Kesimpulan :</p> <p>Penelitian yang dibuat dapat menghasilkan persentase plagiarisme sesuai dengan yang diinginkan, tetapi penelitian ini tidak menjelaskan bahasa pemrograman apa yang digunakan dan database seperti apa yang dipakai.</p>

g. (Jupron & Sutrisno, 2022), *Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat*.

Tabel 2.10 Jurnal ke-7

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Penerapan Algoritma <i>Winnowing</i> untuk Mendeteksi Kemiripan Dua Teks yang Berbeda
2	Jurnal	<p><i>Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat</i></p> <p>p-ISSN 2776-4079 e-ISSN 2808-0033</p>
3	Volume, Nomor, dan Halaman	Volume 2, Nomor 3, Halaman 246-262
4	Bulan dan Tahun	September 2022
5	Penulis	Jupron, Sutrisno
6	Penerbit	-
7	Tujuan Penelitian	<ul style="list-style-type: none"> a. Untuk mendeteksi dan mencegah plagiarisme melalui penggunaan algoritma <i>Winnowing</i>. b. Untuk mengatasi keterbatasan pendeteksian plagiarisme secara manual dengan mengusulkan metode yang dapat membandingkan dokumen dengan ratusan atau bahkan ribuan dokumen lainnya secara efektif.
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Studi literatur b. Pengambilan sampel jurnal c. Metode pengolahan data, seperti <i>preprocessing, Algoritma N-Grams</i> d. Metode algoritma <i>Winnowing</i>
10	Perancangan Sistem	a. Menggunakan Bahasa pemrograman <i>PHP</i>

		b. Menggunakan <i>database SQL</i>
11	Hasil Penelitian	Hasil penelitian ini adalah : a. Halaman <i>login</i> b. Halaman beranda c. Halaman <i>info</i> d. Halaman <i>winnowing</i> e. Halaman tabel f. Aplikasi mampu mendeteksi plagiarisme dari 2 teks dengan baik
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a. Penggunaan algoritma <i>Winnowing</i> yang efektif dalam mendeteksi kemiripan antara dua teks yang berbeda b. Menggunakan algoritma lain yaitu <i>N-Grams</i>
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a. Penjelasan mengenai <i>preprocessing</i> kurang detail b. Tidak disebutkan secara jelas mengenai Batasan ruang lingkup dari penelitian ini.
14	Kesimpulan	Kesimpulan : Penelitian bisa dikatakan berhasil, tetapi masih membingungkan karena kurangnya kedetailan terhadap proses <i>preprocessing</i> nya yang membuat proses terjadinya pengolahan data masih kurang jelas.

h. (Riyani et al., 2019), Jurnal Linguistik Komputasional

Tabel 2.11 Jurnal ke-8

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Penerapan <i>Cosine Similarity</i> dan Pembobotan <i>TF-IDF</i> untuk Mendeteksi Kemiripan Dokumen
2	Jurnal	Jurnal Linguistik Komputasional e-ISSN : 2621-9336
3	Volume, Nomor, dan Halaman	Volume 2, Nomor 1, Halaman 23-27
4	Bulan dan Tahun	Maret 2019
5	Penulis	Ade Riyani, Muhammad Zidny Naf'an, Auliya Burhanuddin
6	Penerbit	<i>Indonesia Association of Computational Linguistics (INACL)</i>

7	Tujuan Penelitian	a. Untuk mendeteksi kemiripan teks dan nilai plagiarisme menggunakan algoritma <i>Cosine Similarity</i> dan pembobotan <i>TF-IDF</i> .
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	a. Metode <i>TF-IDF</i> dan <i>Cosine Similarity</i> b. Metode <i>Preprocessing</i>
10	Perancangan Sistem	a. Data yang digunakan adalah data laporan skripsi mahasiswa dalam bahasa Indonesia b. Aplikasi dapat melakukan <i>preprocessing</i>
11	Hasil Penelitian	Hasil Penelitian ini adalah : a. D1 b. D2 c. D3 d. D4 e. D5 Setiap hasil dokumen memiliki nilai masing masing yang dimana memiliki nilai yang berbeda-beda dan terlihat perbandingan yang di <i>stemming</i> dan yang tidak di <i>stemming</i> .
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a. Memiliki penjelasan <i>text preprocessing</i> yang detail b. Memiliki flowchart proses dari <i>stemming</i> di bagian <i>preprocessing</i> c. Menampilkan hasil antara dilakukannya <i>stemming</i> dan tidak dilakukannya <i>stemming</i> d. Menampilkan tahapan-tahapan proses perhitungan yang detail
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a. Penelitian ini tidak secara rinci menjelaskan program apa saja yang digunakan dan database apa yang dipakai untuk menyimpan data-data laporan skripsi mahasiswa b. Tidak menampilkan aplikasi yang telah jadi. c. Dokumen-dokumen yang digunakan dalam penelitian ini terbatas pada laporan skripsi dalam bahasa Indonesia
14	Kesimpulan	Kesimpulan : Penerapan <i>TF-IDF</i> dan <i>Cosine Similarity</i> dalam penelitian ini di jelaskan dengan baik, penelitian ini memberikan hasil

		dalam bentuk tabel yang mudah dimengerti. Walaupun begitu penjelasan dalam masalah program yang dipakai masih belum menjelaskan dengan baik.
--	--	--

i. (Apridiansyah et al., 2022), Jurnal Media Infotama.

Tabel 2.12 Jurnal ke-9

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Penerapan Fungsi Metode <i>Rolling Hash</i> Pada Algoritma <i>Winnowing</i> Untuk Mendeteksi Kemiripan Teks Abstrak Berbasis <i>Web</i>
2	Jurnal	Jurnal Media Infotama p-ISSN : 1858-2680 e-ISSN : 2723-4673
3	Volume, Nomor, dan Halaman	Volume 18, Nomor 1, Halaman 128-133
4	Bulan dan Tahun	April 2022
5	Penulis	Yovi Apridiansyah, Ardi Wijaya, Ari Purjiawan
6	Penerbit	Universitas Dehasen Bengkulu
7	Tujuan Penelitian	a. Untuk menerapkan metode <i>rolling hash</i> pada algoritma <i>winnowing</i> guna mendeteksi plagiarisme dalam bahasa Indonesia. b. Untuk mengidentifikasi dan mencegah plagiarisme dalam dokumen teks.
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	a. Metode algoritma <i>winnowing</i> b. Metode <i>N-Gram</i> c. Metode <i>Rolling Hash</i> d. Metode Analisa, <i>design, code, test</i> e. Metode <i>jaccard Coeficient</i> f. Metode <i>model incremental</i>
10	Perancangan Sistem	a. Menggunakan metode <i>model incremental</i> . b. Bahasa matlab
11	Hasil Penelitian	Hasil penelitian ini adalah : a. Tampilan <i>Login</i> b. Tampilan Utama c. Tampilan Proses Perbandingan Kemiripan Abstrak d. Tampilan Hasil Proses e. Hasil yang didapatkan adalah 50,06%

12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a. Penerapan metode <i>rolling hash</i> pada algoritma <i>winnowing</i> , yang merupakan metode efektif dalam mendeteksi kemiripan teks dan plagiarisme b. Penggunaan <i>N-gram</i> dalam mengolah data menjadi informasi,
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a. Tidak menjelaskan secara detail program apa saja yang dipakai b. Tidak adanya penjelasan <i>database</i> secara detail c. Perhitungan algoritma <i>winnowing</i> tidak dijelaskan dengan baik
14	Kesimpulan	Kesimpulan : Penelitian yang menghasilkan sebuah aplikasi web plagiarisme ini telah menghasilkan metode yang menarik, penelitian ini telah berhasil menghasilkan plagiat yang menunjukkan hasil 50,06%, tetapi dalam penelitian ini tidak menjelaskan secara detail apa saja yang diperlukan dalam menciptakan program aplikasinya walaupun berhasil mendapatkan hasilnya tapi penjelasan dalam perancangan system programnya tidak terlalu detail.

j. (Ariyanti et al., 2021), JASISFO (Jurnal Sistem Informasi).

Tabel 2.13 Jurnal ke-10

No	Data Jurnal/ Makalah	Keterangan
1	Judul	Perancangan Aplikasi Pendeteksi Plagiarisme Teks Menggunakan Metode <i>Rational Unified Process</i> Pada Aplikasi <i>E-Learning</i>
2	Jurnal	JASISFO (Jurnal Sistem Informasi) p-ISSN 2721-1614 e-ISSN 2722-3418
3	Volume, Nomor, dan Halaman	Volume 2, Nomor 1, Halaman 141-149
4	Bulan dan Tahun	Maret 2021
5	Penulis	Indri Ariyanti, M Aris Ganiardi, Delta Khairunnisa
6	Penerbit	-

7	Tujuan Penelitian	<ul style="list-style-type: none"> a. Untuk membantu dosen dalam mengkoreksi tugas mahasiswa dengan memudahkan pendeteksian kesamaan jawaban yang dikirim oleh mahasiswa. b. Untuk merancang suatu fitur aplikasi e-learning yang mampu mendeteksi plagiarisme pada dokumen teks tugas mahasiswa.
8	Lokasi dan Objek Penelitian	Lokasi penelitian ini adalah di kota Palembang. Objek penelitian ini adalah aplikasi pendeteksi plagiarisme teks pada dokumen menggunakan metode <i>Rational Unified Process (RUP)</i> .
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Metode <i>Rational Unified Process (RUP)</i> b. Metode Pengumpulan Data, seperti survei, observasi ke perpustakaan, wawancara dengan penulis artikel, dosen, dan mahasiswa yang sering membuat tugas akhir. c. Metode algoritma <i>n-gram</i>
10	Perancangan Sistem	<ul style="list-style-type: none"> a. Menggunakan <i>database MySQL</i> b. Menggunakan bahasa pemrograman Java. c. Rancangan modul dapat berbentuk algoritma, notasi <i>UML</i>, dan <i>pseudo-code</i>. d. Menggunakan metode <i>Rational Unified Process (RUP)</i>
11	Hasil Penelitian	<p>Hasil penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Analisis Pendeteksian Plagiarisme b. Deskripsi Umum Perangkat Lunak c. <i>Diagram Use Case</i> d. Diagram Kelas e. Rancangan Antarmuka
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Menggunakan metode <i>Rational Unified Process (RUP)</i>. b. Melibatkan pengumpulan data seperti studi literatur, wawancara, survei, dan observasi, yang menghasilkan informasi-informasi detail. c. Menggunakan algoritma <i>n-gram</i>.
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Tidak ada informasi atau tahap-tahap dalam pengerjaan program. b. Tidak ada informasi lanjutan tentang hasil uji coba program.

		c. Tidak ada informasi bagaimana aplikasi ini akan diimplementasikan ke <i>e-learning</i> .
14	Kesimpulan	Kesimpulan : Penelitian ini bisa dijadikan ide untuk meningkatkan nilai dari sebuah <i>e-learning</i> dengan adanya <i>system</i> plagiarisme yang di implementasikan ke sebuah <i>e-learning</i> maka tugas-tugas yang kemungkinan besar hasil dari contekan melalui teman yang lain dapat ketahuan, tetapi dalam penelitian ini masih kurang menjelaskan bagaimana perhitungan lanjutan dari algoritma yang dipakai, tahap-tahap pembuatan program, dan tentunya bagaimana mengimplementasikan ke sebuah <i>e-learning</i> .

k. (Ahuja et al., 2020), *Arabian Journal for Science and Engineering*.

Tabel 2.14 Jurnal ke-11

No	Data Jurnal/ Makalah	Keterangan
1	Judul	<i>A New Hybrid Technique for Detection of Plagiarisme from Text Documents</i>
2	Jurnal	<i>Arabian Journal for Science and Engineering</i> p-ISSN 2193-567X e-ISSN 2191-4281
3	Volume, Nomor, dan Halaman	Volume 45, Nomor 12, Halaman 9939-9952
4	Bulan dan Tahun	April 2020
5	Penulis	<i>Lovepreet Ahuja, Vishal Gupta, Rohit Kumar</i>
6	Penerbit	<i>Springer Link</i>
7	Tujuan Penelitian	a. Untuk mengembangkan sistem pendeteksian plagiarisme lintas bahasa untuk dokumen-dokumen dalam bahasa Arab. b. Untuk menciptakan deteksi plagiarisme yang Kompleks.
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	d. Metode <i>preprocessing</i> e. Metode pendekatan deteksi plagiarisme yang menggunakan

		pengetahuan semantik dan fitur <i>linguistic</i> .
		f. Metode <i>A New Hybrid Technique</i>
10	Perancangan Sistem	a. Database leksikal <i>WordNet</i> b. Algoritma <i>preprocessing</i> c. Bahasa pemrograman <i>python</i>
11	Hasil Penelitian	Hasil penelitian ini adalah : a. Metrik Evaluasi b. Kumpulan dataata yang digunakan untuk eksperimen c. Eksperimen dan diskusi d. Perbandingan dengan sistem PAN-PC-11 e. Validasi hasil
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a. Metodologi yang solid b. Menjelaskan perhitungan dengan detail c. Memberikan code yang dipakai
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a. Aplikasi yang telah dibuat tidak dijelaskan secara detail b. <i>Database</i> yang dipakai tidak dijelaskan secara detail
14	Kesimpulan	Kesimpulan : Metode yang dipakai dalam penelitian ini dapat menjadi terobosan baru, metode ini dapat mendeteksi kata-kata plagiarisme secara kompleks, tetapi masih kurang penjelasan dalam pembuatan aplikasi walaupun telah ditampilkan program dalam penelitian ini, pengimplementasian program tersebut masih belum jelas.

1. (Muangprathub et al., 2021), *Journal of Applied Mathematics*

Tabel 2.15 Jurnal ke-12

No	Data Jurnal/ Makalah	Keterangan
1	Judul	<i>Document Plagiarisme Detection Using a New Concept Similarity in Formal Concept Analysis</i>
2	Jurnal	<i>Journal of Applied Mathematics</i>
3	Volume, Nomor, dan Halaman	Volume 2021, -, Halaman 1-10 p-ISSN 1110-757X e-ISSN 1687-0042
4	Bulan dan Tahun	March 2021

5	Penulis	Jirapond Muangprathub, Siriwan Kajornkasirat, Apirat Wanichsombat
6	Penerbit	Hindawi
7	Tujuan Penelitian	<ol style="list-style-type: none"> Mengembangkan algoritma untuk membangun struktur informasi dalam deteksi plagiarisme dokumen. Menyediakan pendekatan baru dalam deteksi plagiarisme dokumen yang dapat meningkatkan akurasi dan efisiensi dalam proses identifikasi dokumen yang plagiarisme. Menyediakan sistem deteksi plagiarisme dokumen yang dapat digunakan untuk berbagai jenis dokumen.
8	Lokasi dan Objek Penelitian	Lokasi di <i>Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus</i> , Surat Thani 84000, Thailand dan objek penelitiannya adalah pengembangan algoritma deteksi plagiarisme dokumen menggunakan pendekatan konsep formal.
9	Metode yang digunakan	<ol style="list-style-type: none"> Metode analisis <i>formal concept</i> Pengukuran Kesamaan Konsep
10	Perancangan Sistem	<ol style="list-style-type: none"> Bahasa pemrograman <i>php</i> Database <i>mysql</i> Aplikasi <i>Online</i>
11	Hasil Penelitian	Hasil penelitian ini adalah : <ol style="list-style-type: none"> Tampilan Utama Sistem Sistem ini mampu mendeteksi tingkat plagiarisme dalam dokumen dengan akurasi rata-rata sebesar 94.01%
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : <ol style="list-style-type: none"> Implementasi Sistem Berbasis <i>Web</i> Tingkat Akurasi Tinggi Menjelaskan perhitungan dengan detail
13	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : <ol style="list-style-type: none"> Keterbatasan Pengujian pada Bahasa Thailand Keterbatasan Studi Kasus
14	Kesimpulan	Kesimpulan : Penelitian ini menunjukkan bahasa metode pendekatan <i>FCA</i> dapat digunakan dengan baik dalam menganalisis struktur informasi dokumen, penelitian ini juga dapat mendukung deteksi plagiarisme dengan tingkat akurasi yang tinggi hingga 94.01%, tetapi proses penjelasan dan pembuatan aplikasinya masih

		membingungkan karena kurang detail dalam penjelasannya
--	--	--

m. (Antonius et al., 2023), *International Journal of Advanced Computer Science and Applications*

Tabel 2.16 Jurnal ke-13

No	Data Jurnal/ Makalah	Keterangan
1	Judul	<i>Enhanced Plagiarisme Detection Through Advanced Natural Language Processing and E-BERT Framework of the Smith-Waterman Algorithm</i>
2	Jurnal	(IJACSA) <i>International Journal of Advanced Computer Science and Applications</i>
3	Volume, Nomor, dan Halaman	Volume 14, Nomor 9, Halaman 408-416 p-ISSN 2158-107X e-ISSN 2156-5570
4	Bulan dan Tahun	2023
5	Penulis	<i>Franciskus Antonius, Myagmarsuren Orosoo, Aanandha Saravanan K, Indrajit Patra, Prema S</i>
6	Penerbit	<i>The Science and Information (SAI) Organization</i>
7	Tujuan Penelitian	a. Untuk meningkatkan akurasi deteksi plagiarisme dengan algoritma <i>Smith-Waterman</i> dan teknik kamus bahasa Inggris-Spanyol. b. Untuk mengatasi keterbatasan sistem deteksi plagiarisme yang sudah ada dengan memperbaiki akurasi dan efektivitas deteksi plagiarisme.
8	Lokasi dan Objek Penelitian	Objek penelitian ini adalah pengembangan metode deteksi plagiarisme yang lebih canggih menggunakan pendekatan <i>Natural Language Processing (NLP)</i> dan <i>framework E-BERT</i> dari algoritma <i>Smith-Waterman</i> .
9	Metode yang digunakan	a. Metode algoritma <i>Smith-Waterman</i> b. Metode pendekatan <i>Natural Language Processing (NLP)</i> c. Metode <i>preprocessing</i>
10	Perancangan Sistem	a. Bahasa pemrograman <i>python</i>
11	Hasil Penelitian	a. Tampilan hasil kinerja sistem deteksi plagiarisme dalam bentuk metrik

		<p>seperti akurasi, presisi, <i>recall</i>, dan <i>F1-score</i></p> <p>b. Hasil akurasi yang didapatkan mencapai 99,5%</p>
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <p>a. Penggunaan pendekatan yang canggih seperti <i>Natural Language Processing (NLP)</i> dan <i>framework E-BERT</i> untuk deteksi plagiarisme.</p> <p>b. Implementasi algoritma <i>Smith-Waterman</i> yang dapat meningkatkan akurasi dalam identifikasi plagiarisme.</p> <p>c. Penggunaan metrik kinerja yang komprehensif seperti akurasi, presisi, <i>recall</i>, <i>F1-score</i>, dan <i>Area Under the Curve (AUC)</i> untuk evaluasi sistem deteksi plagiarisme.</p>
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <p>a. Tidak disebutkan secara spesifik mengenai penggunaan database atau lingkungan pengujian yang lebih rinci.</p> <p>b. Tidak dijelaskan data yang dipakai data apa.</p>
14	Kesimpulan	<p>Kesimpulan :</p> <p>Penelitian ini telah berhasil mengimplementasikan teknik algoritma <i>Smith-Waterman</i> dengan baik, sehingga mencapai tingkat akurasi yang sangat tinggi sebesar 99,5%, tetapi dalam penelitian ini tidak menjelaskan secara detail aplikasi yang sudah jadi, tampilan-tampilan dari aplikasi juga tidak disebutkan, walaupun memakai bahasa pemrograman <i>python</i> tetapi pengimplementasian program tersebut masih kurang detail</p>

n. (Chang et al., 2021), *Information Retrieval Journal (Discover Computing)*

Tabel 2.17 Jurnal ke-14

No	Data Jurnal/ Makalah	Keterangan
1	Judul	<i>Using word semantic concepts for plagiarism detection in text documents</i>
2	Jurnal	<i>Information Retrieval Journal (Discover Computing)</i>
3	Volume, Nomor, dan Halaman	Volume 24, Volume 4-5, Halaman 298–321

		p-ISSN 2948-2984 e-ISSN 2948-2992
4	Bulan dan Tahun	July 2021
5	Penulis	<i>Chia-Yang Chang, Shie-Jue Lee, Chih-Hung Wu, Chih-Feng Liu, Ching-Kuan Liu</i>
6	Penerbit	Springer Link
7	Tujuan Penelitian	<p>a. Untuk mengembangkan metode deteksi plagiarisme yang memperhitungkan hubungan semantik antara kata-kata dalam representasi dokumen</p> <p>b. Untuk mengembangkan pendekatan dua tahap, yaitu tahap penyaringan dan tahap identifikasi, untuk deteksi plagiarisme yang efektif.</p>
8	Lokasi dan Objek Penelitian	Penelitian ini berfokus pada pengembangan metode deteksi plagiarisme yang memperhitungkan hubungan semantik antara kata-kata dalam representasi dokumen
9	Metode yang digunakan	<p>a. Metode <i>spherical K-means</i></p> <p>b. Metode <i>two-phase matching strategy</i></p> <p>c. Metode pengambilan dan penimbangan bidang</p> <p>d. Metode <i>fuzzy semantic-based string similarity approach</i></p> <p>e. Metode <i>MultiLayer Self-Organizing Map (MLSOM)</i></p> <p>f. Metode <i>MultiLevel Matching (MLM)</i></p> <p>g. Metode <i>Semantic compression</i></p> <p>h. Metode <i>Word2vec</i></p>
10	Perancangan Sistem	a. Datasets yang digunakan berasal dari <i>HTML_CityUI dataset, PAN Plagiarisme Corpus 2010, PAN-PC-10</i>
11	Hasil Penelitian	<p>Hasil penelitian ini adalah :</p> <p>a. <i>Performance measures</i></p> <p>b. Metode menggunakan <i>Word2vec</i> untuk mengubah kata-kata menjadi vektor kata dan mengelompokkannya ke dalam konsep semantik terbukti efektif.</p> <p>c. Hasil eksperimen menunjukkan bahwa pendekatan ini mampu meningkatkan presisi, <i>recall, granularity</i>, dan skor deteksi plagiarisme (<i>PlagDet</i>) dibandingkan dengan metode lain seperti <i>MLMH</i>.</p>
12	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah :

		<p>a. Penggunaan pendekatan <i>Word2vec</i> untuk mengubah kata-kata menjadi vektor kata dan mengelompokkannya ke dalam konsep semantik, yang terbukti efektif dalam meningkatkan deteksi plagiarisme.</p> <p>b. Eksperimen yang dilakukan menunjukkan peningkatan signifikan dalam presisi, <i>recall</i>, <i>granularity</i>, dan skor deteksi plagiarisme (<i>PlagDet</i>) dibandingkan dengan metode tradisional seperti <i>MLMH</i>.</p>
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <p>a. Ketergantungan pada kualitas data pelatihan</p> <p>b. Keterbatasan generalisasi</p> <p>c. Memiliki tantangan dalam menangani perubahan gaya penelitian</p>
14	Kesimpulan	<p>Kesimpulan :</p> <p>Pendekatan menggunakan <i>Word2vec</i> untuk mengubah kata-kata menjadi vektor kata dan mengelompokkannya ke dalam konsep semantik telah terbukti efektif dalam meningkatkan deteksi plagiarisme dalam dokumen teks. Meskipun terdapat kelemahan seperti keterbatasan dalam representasi kata-kata dan ketergantungan pada kualitas data pelatihan, penelitian ini memberikan wawasan yang berharga tentang potensi penggunaan <i>Word2vec</i> dalam konteks deteksi plagiarisme.</p>

o. (El-Rashidy et al., 2024), *Multimedia Tools and Applications*

Tabel 2.18 Jurnal ke-15

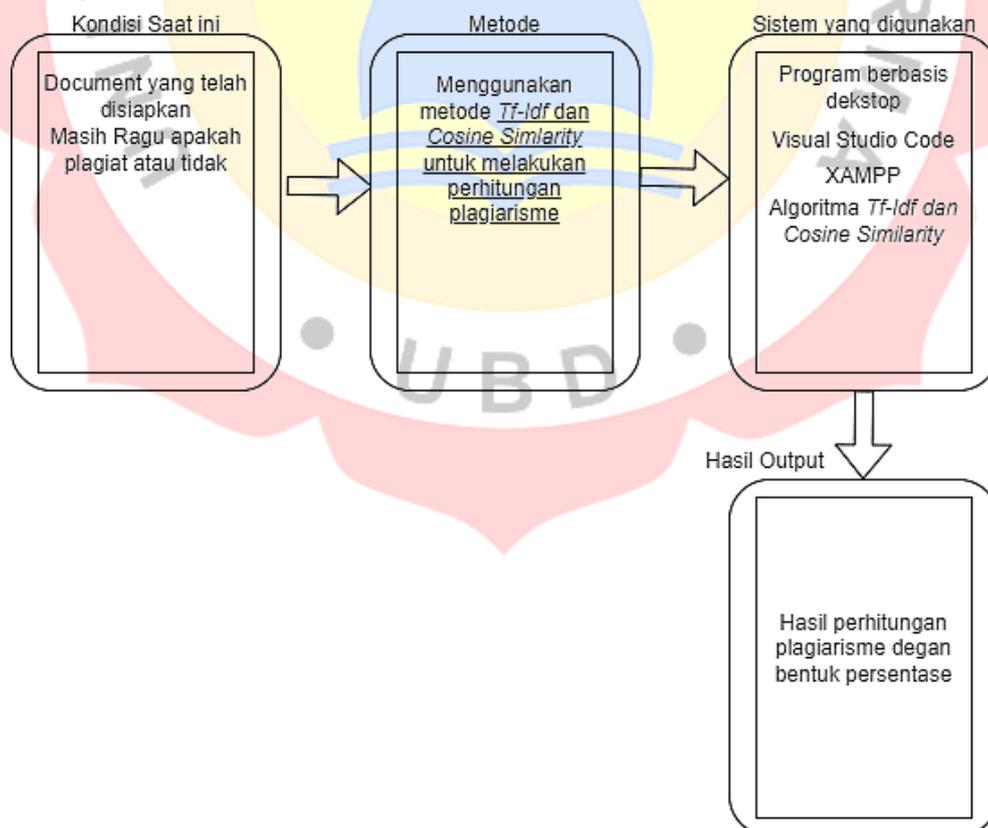
No	Data Jurnal/ Makalah	Keterangan
1	Judul	<i>An efective text plagiarisme detection system based on feature selection and SVM techniques</i>
2	Jurnal	<i>Multimedia Tools and Applications</i>
3	Volume, Nomor, dan Halaman	Volume 83, Nomor 1, Halaman 2609-2646 p-ISSN 1380-7501 e-ISSN 1573-7721
4	Bulan dan Tahun	May 2023

5	Penulis	<i>Mohamed A. El-Rashidy, Ramy G. Mohamed, Nawal A. El-Fishawy, Marwa A. ShoumanI</i>
6	Penerbit	<i>Link Springer</i>
7	Tujuan Penelitian	<ul style="list-style-type: none"> a. Mengusulkan sistem deteksi plagiarisme baru yang menangani semua jenis plagiarisme, seperti plagiarisme leksikal, sintaksis, dan semantic b. untuk meningkatkan akurasi deteksi plagiarisme dengan menggunakan teknik pemrosesan bahasa alami dan pemilihan fitur yang tepat.
8	Lokasi dan Objek Penelitian	-
9	Metode yang digunakan	<ul style="list-style-type: none"> a. Metode <i>preprocessing</i> b. Metode teknik <i>SVM</i> c. Metode pemilihan fitur d. Algoritma <i>chi-square</i>
10	Perancangan Sistem	<ul style="list-style-type: none"> a. Menggunakan metode <i>preprocessing</i> b. Memiliki Fitur <i>Fuzzy similarity</i>
11	Hasil Penelitian	<p>Hasil penelitian :</p> <ul style="list-style-type: none"> a. Sistem menunjukkan kinerja yang superior dalam mengatasi kasus-kasus plagiarisme acak dan berbagai jenis plagiarisme teks lainnya. b. Sistem ini mencapai tingkat <i>PlagDet</i> sebesar 89.12% dan <i>F-measure</i> sebesar 89.34% pada dokumen uji <i>obfuscation</i> acak dalam dataset <i>PAN 2013</i>
12	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Pemilihan fitur menggunakan algoritma <i>Chi-square</i> untuk mengekstrak fitur paling diskriminatif untuk jenis plagiarisme teks yang berbeda. b. Penggunaan algoritma klasifikasi <i>SVM</i> untuk membangun persamaan hiperplane dari 34 fitur, menghasilkan kinerja yang baik tanpa perlu melakukan eksperimen ekstensif untuk menemukan nilai koefisien bobot terbaik. c. Perhitungan yang detail
13	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a. Ketergantungan pada dataset <i>PAN 2012</i> dan <i>PAN 2013</i> untuk melatih algoritma <i>SVM</i>, yang mungkin membatasi generalisasi sistem

		terhadap kasus plagiarisme teks yang lebih luas. b. Tidak dijelaskan pemakaian <i>database</i> apa yang digunakan.
14	Kesimpulan	Kesimpulan : Penggunaan metode algoritma <i>chi-square</i> dan <i>svm</i> menunjukkan hasil yang baik, tetapi pengimplementasian <i>codenya</i> masih kurang detail, seperti bagaimana cara <i>codenya</i> berkerja dan bagaimana <i>code</i> itu dibentuk sedemikian rupa menjadi aplikasi.

Dari hasil tinjauan jurnal diatas maka penelitian ini akan menggunakan metode *td-idf* dan *Cosine Similarity* karena dalam tinjauan jurnal diatas metode-metode dapat dikatakan sebagai metode yang bagus dalam mendeteksi plagiarisme, dan dalam tinjauan jurnal diatas metode *td-idf* dan *Cosine Similarity* adalah metode yang sering digunakan dalam penggunaan aplikasi plagiarisme.

2.6 Kerangka Pemikiran

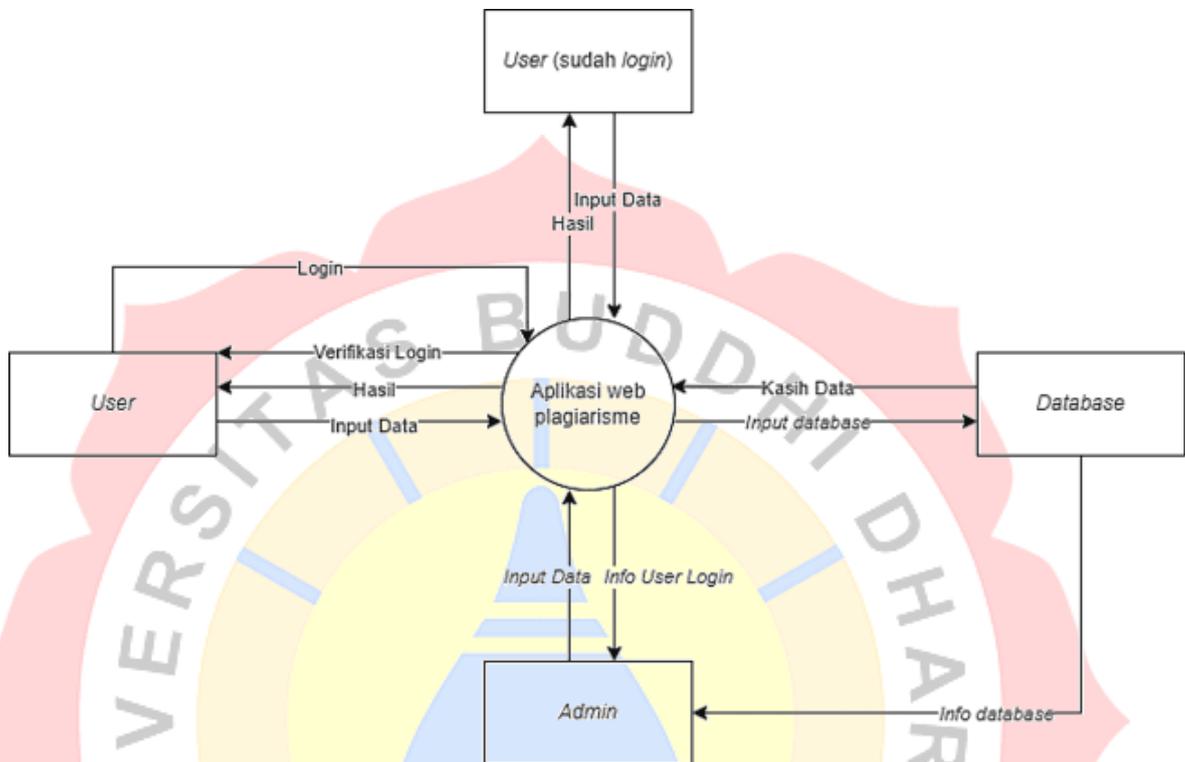


Gambar 2.14 Kerangka Pemikiran

BAB III

METODOLOGI PENELITIAN

3.1 Data Flow Diagram (DFD)



Gambar 3.1 Data Flow Diagram

3.2 Analisa

3.2.1 Masukan

Proses masukan akan terdiri dari 3 menu yang dimana menu masukan user biasa, menu user yang sudah login, menu admin. Proses masukan tersebut akan dimulai dari proses pemasukan data melalui menu masukan file yang sudah dibentuk sedemikian rupa agar dapat menerima data berupa pdf.

File-file tersebut akan diproses terlebih dahulu agar proses lanjutan bisa dilakukan dengan baik. Menu admin adalah menu yang dikhususkan untuk admin dalam memasukkan data uji kedalam database, admin memiliki hak untuk memasukkan data-data yang akan dijadikan sebagai data uji nantinya.

3.2.2 Proses

Setelah terjadinya proses memasukkan data, proses utama akan terjadi yang dimana proses pertama yang dimulai adalah proses *preprocessing* proses *preprocessing* dapat dikatakan sebagai proses pembuka agar proses perhitungan berhasil dilakukan. Proses *preprocessing* dapat dikatakan akan memakan waktu yang lumayan lama karena tahap ini terjadi proses pembersihan data dari *noise*, *number*, ataupun data yang error.

Setelah proses *preprocessing* sudah siap, dilakukanlah proses utama yaitu proses perhitungan yang menggunakan metode *TF-IDF* dan *Cosine Similarity*, dua metode ini akan menjadi inti utama dalam memeriksa dan mengecek plagiat dari dokumen yang sudah diproses. Proses akan terlebih dahulu dilakukan di *TF-IDF* untuk menghitung bobot dari sebuah kata, proses ini adalah proses awal yang dimana hasil dari *TF-IDF* akan dipakai untuk proses *Cosine Similarity*, inilah proses utama pengecekan plagiat yang menggunakan rumus khusus untuk mendeteksi plagiat setiap kata dari dokumen data uji dengan dokumen *user*.

3.2.3 Keluaran

Proses keluaran adalah proses hasil yang telah didapatkan oleh perhitungan dari metode *TF-IDF* dan *Cosine Similarity*. Hasil dari perhitungan akan di tampilkan dalam bentuk persentase. Persentase tersebut adalah hasil dari plagiarisme yang paling tinggi dari dokumen-dokumen yang ada dalam *database* yang dijadikan sebagai data uji.

3.2.4 Data

Data yang dipakai akan dimasukkan ke dalam database dari *menu admin*, data-data tersebut dikumpulkan agar data *user* dapat dibandingkan dengan data uji. Data-data tersebut dikumpulkan ke dalam *database* dimana terdapat *database* khusus dalam

mysql yang dibuat secara khusus untuk menampung semua data yang didapatkan dari 1 data.

3.3 Analisa Masalah

Dari analisa yang dilakukan, terdapat banyak masalah yang sedang terjadi di dunia mengenai plagiarisme, seperti :

1. Banyak aksi plagiarisme yang terjadi di dunia ini.
2. Aksi-aksi plagiarisme telah menjadi permasalahan yang terpenting dalam dunia pendidikan.
3. Plagiarisme yang sudah melekat erat di dunia ini.

3.4 Identifikasi kebutuhan sistem

Dari analisa masalah yang dilakukan, ditemukanlah sebuah cara yang bisa dilakukan untuk memberantas plagiarisme seperti aplikasi plagiarisme, maka penelitian ini membuat aplikasi yang dapat memeriksa plagiat dengan sistem seperti :

1. Aplikasi harus dapat menghitung plagiat suatu dokumen
2. Hasil perhitungan di tampilkan
3. Kata-kata yang terkena plagiat ditampilkan
4. Memiliki *history* plagiat bagi user yang sudah login

3.5 Metode *TF-IDF* dan *Cosine Similarity*

Dalam sebuah penelitian. Perhitungan adalah yang paling utama, perhitungan - perhitungan ini yang akan menghasilkan persentase plagiarisme, perhitungan manual dari *TF-IDF* dan *Cosine Similarity* memiliki beberapa tahap yang harus dikerjakan terlebih dahulu, berbeda dengan menggunakan algoritmanya yang menggunakan fungsi dari *pythonnya* sendiri.

Contoh perhitungan manual dari *TF-IDF* dan *Cosine Similarity* bisa dilihat dari gambar dan perhitungan berikut ini:

1. Contoh soal

U : Pada era globalisasi saat ini, perkembangan pola hidup Masyarakat

D1: Plagiarisme tindak langgar momok ilmu tahu laku plagiarisme

D2: lapor teliti kembang fakultas ilmu dana boptn tahun anggar

D3: aplikasi terap algoritma cocok string dokumen

2. Tf

Tabel 3.1 Tabel tf

Tokens	u	d1	d2	d3
era	1	0	0	0
globalisasi	1	0	0	0
ini	1	0	0	0
kembang	1	0	1	0
pola	1	0	0	0
hidup	1	0	0	0
masyarakat	1	0	0	0
plagiarisme	0	1	0	0
tindak	0	1	0	0
langgar	0	1	0	0
momok	0	1	0	0
ilmu	0	1	1	0
tahu	0	1	0	0
laku	0	1	0	0
lapor	0	0	1	0
teliti	0	0	1	0
fakultas	0	0	1	0
dana	0	0	1	0
boptn	0	0	1	0
tahun	0	0	1	0
anggar	0	0	1	0
aplikasi	0	0	0	1
terap	0	0	0	1
algoritma	0	0	0	1
cocok	0	0	0	1
string	0	0	0	1
dokumen	0	0	0	1

3. Df

Tabel 3.2 Tabel *df*

Tokens	u	d1	d2	d3	df
era	1	0	0	0	1
globalisasi	1	0	0	0	1
ini	1	0	0	0	1
kembang	1	0	1	0	2
pola	1	0	0	0	1
hidup	1	0	0	0	1
masyarakat	1	0	0	0	1
plagiarisme	0	1	0	0	1
tindak	0	1	0	0	1
langgar	0	1	0	0	1
momok	0	1	0	0	1
ilmu	0	1	1	0	2
tahu	0	1	0	0	1
laku	0	1	0	0	1
lapor	0	0	1	0	1
teliti	0	0	1	0	1
fakultas	0	0	1	0	1
dana	0	0	1	0	1
boptn	0	0	1	0	1
tahun	0	0	1	0	1
anggar	0	0	1	0	1
aplikasi	0	0	0	1	1
terap	0	0	0	1	1
algoritma	0	0	0	1	1
cocok	0	0	0	1	1
string	0	0	0	1	1
dokumen	0	0	0	1	1

Setelah tabel *df*nya ketemu saatnya melakukan proses perhitungan *idf*, dengan

rumus :

$$Idf = \text{LOG}\left(\frac{N}{df}\right)$$

4. Idf

Tabel 3.3 Tabel *idf*

Tokens	u	d1	d2	d3	df	idf
era	1	0	0	0	1	0.60206
globalisasi	1	0	0	0	1	0.60206

ini	1	0	0	0	1	0.60206
kembang	1	0	1	0	2	0.30103
pola	1	0	0	0	1	0.60206
hidup	1	0	0	0	1	0.60206
masyarakat	1	0	0	0	1	0.60206
plagiarisme	0	1	0	0	1	0.60206
tindak	0	1	0	0	1	0.60206
langgar	0	1	0	0	1	0.60206
momok	0	1	0	0	1	0.60206
ilmu	0	1	1	0	2	0.30103
tahu	0	1	0	0	1	0.60206
laku	0	1	0	0	1	0.60206
lapor	0	0	1	0	1	0.60206
teliti	0	0	1	0	1	0.60206
fakultas	0	0	1	0	1	0.60206
dana	0	0	1	0	1	0.60206
boptn	0	0	1	0	1	0.60206
tahun	0	0	1	0	1	0.60206
anggar	0	0	1	0	1	0.60206
aplikasi	0	0	0	1	1	0.60206
terap	0	0	0	1	1	0.60206
algoritma	0	0	0	1	1	0.60206
cocok	0	0	0	1	1	0.60206
string	0	0	0	1	1	0.60206
dokumen	0	0	0	1	1	0.60206

5. *TF-IDF*

Tabel 3.4 Tabel *TF-IDF*

Tokens	u	d1	d2	d3
era	0.60206	0	0	0
globalisasi	0.60206	0	0	0
ini	0.60206	0	0	0
kembang	0.30103	0	0.30103	0
pola	0.60206	0	0	0
hidup	0.60206	0	0	0
masyarakat	0.60206	0	0	0
plagiarisme	0	0.60206	0	0
tindak	0	0.60206	0	0
langgar	0	0.60206	0	0
momok	0	0.60206	0	0
ilmu	0	0.30103	0.30103	0
tahu	0	0.60206	0	0
laku	0	0.60206	0	0
lapor	0	0	0.60206	0

teliti	0	0	0.60206	0
fakultas	0	0	0.60206	0
dana	0	0	0.60206	0
boptn	0	0	0.60206	0
tahun	0	0	0.60206	0
anggar	0	0	0.60206	0
aplikasi	0	0	0	0.60206
terap	0	0	0	0.60206
algoritma	0	0	0	0.60206
cocok	0	0	0	0.60206
string	0	0	0	0.60206
dokumen	0	0	0	0.60206

6. Cosine Similarity

proses perhitungan *Cosine Similarity* dapat dijalankan dengan rumus:

$$SV = \frac{AB}{||A|| ||B||} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

U = Pada era globalisasi ini, perkembangan pola hidup Masyarakat

Tabel 3.5 Tabel Hasil

d1	atas	0
	bawah	2.26547646
hasil		0
d2	atas	0.09061906
	bawah	2.48170512
hasil		4%
d3	atas	0
	bawah	2.21970454
hasil		0

Hasil:

$$D1 = 0$$

$$D2 = 4\%$$

$$D3 = 0$$

Sudah didapatkan hasil perhitungan, perhitungan ini hanya sebuah contoh yang diambil dari sebagian dataset dan data uji. Perhitungan menggunakan sebagian ini

dikarenakan data yang sangat banyak, oleh karena itu perhitungan hanya akan mengambil sebagian data saja.

3.6 Requirement Elicitation

Tabel 3.6 Requirement Elicitation

No	Analisa Kebutuhan	Keterangan
1	User bisa melakukan pengecekan tanpa ada pembatasan	✓
2	<i>Drag and drop box menu</i>	✓
3	Kasih tampilan proses bar pengecekan	x
4	Kasih judul dan kasih link	✓
5	<i>Pdf</i> dikasih user di tampilin sesuai dokumen yang terplagiat dan diwarnai	x
6	Di hasil dokumen user akan berbentuk <i>pdf</i>	✓
7	Memberikan waktu progress di halaman hasil	✓

Tabel 3.7 Requirement Elicitation

No	Analisa Kebutuhan	Keterangan
1	<i>Runtime</i> dipercepat	✓
2	Tampilan dirapihin	✓
3	Hasil memakai grafika lingkaran	x
4	<i>Font Family</i> diganti	✓

3.7 Jadwal Penelitian

Tabel 3. 8 Tabel Jadwal Penelitian

No	Kegiatan	2024															
		Maret				April				Mei				Juni			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data																
2	Analisa Data																

