

IMPLEMENTASI TEXT MINING DALAM MENGIDENTIFIKASI SIMILARITY JUDUL PENELITIAN

Robi Yanto, Alfiarini, Hendra Di Kesuma, Syafri Afrudi, Juminovario, Ahmadi

1,2,6) Sistem Informasi, STMIK Bina Nusantara Jaya Lubuk Linggau, Lubuklinggau, Indonesia

3) Sistem Informasi, Ilmu Komputer, Universitas Indo Global Mandiri, Palembang, Indonesia

4,5) Ilmu Komputer, Teknik, Universitas Musi Rawas, Musi Rawas, Indonesia

Article Info

Article history:

Received: 10 Maret 2025

Revised: 17 Maret 2025

Accepted: 20 Maret 2025

ABSTRACT

Abstrak

Plagiarisme adalah sebuah tindakan mengambil sebagian atau karya milik orang lain dan menjadikannya sebagai suatu karya asli milik sendiri tanpa menyantumkan sumber secara tepat. Plagiarisme terjadi disegala bidang kehidupan salah satunya plagiarisme judul skripsi. Dalam mengatasi plagiarisme judul perlu dibuatkan sebuah sistem yang menggunakan text mining dalam pengolahan kata untuk mengetahui presentase kesamaan judul, agar mengurangi tingkat plagiarisme. Algoritma winnowing adalah suatu teknik yang digunakan untuk mendeteksi kesamaan teks menggunakan fungsi hashing yang bekerja dengan teknik fingerprinting. Hasil pengolahan dapat mengetahui presentase kesamaan antara judul yang diajukan dengan judul yang sudah ada dengan tingkat akurasi 100 % antara teks dan pola pencarian. Sselain itu dengan perhitungan Jaccard Similarity memberikan keunggulan dalam menentukan tingkat similarity yang lebih tinggi dan membantu dalam pencarian literatur yang relevan. Dengan teknik tersebut diharapkan dapat membantu pihak program studi dalam melakukan pemeriksaan judul penelitian dan meningkatkan akurasi, efektifitas dan skalabilitas serta mengurangi tingkat plagiarisme.

Kata Kunci: Winnowing, Text Mining, Similarity, Judul Penelitian

Abstract

Plagiarism is an act of taking part or work of others and making it a original work of his own without including the source appropriately. Plagiarism occurs in all fields of life, one of which is plagiarism thesis title. In overcoming plagiarism the title needs to be made a system that uses text mining in word processing to find out the percentage of similarity in order, in order to reduce the level of plagiarism. Winnowing algorithm is a technique used to detect the similarity of text using the hashing function that works with fingerprinting techniques. Processing results can find out the percentage of similarities between the proposed titles with the existing title with a 100 % accuracy level between text and search patterns. In addition, the calculation of Jaccard Similarity provides excellence in determining the higher Similarity level and helps in searching for relevant literature. With this technique, it is expected to be able to assist the study program in conducting research titles and increasing accuracy, effectiveness and scalability and reducing the level of plagiarism.

Keywords: Plagiarism, Winnowing, Text Mining, Similarity, Research Title

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail : pakrobibnj@gmail.com

1. PENDAHULUAN

Pada perguruan tinggi, mewajibkan mahasiswa membuat sebuah karya ilmiah yaitu skripsi sebagai syarat untuk mendapatkan gelar sarjana. Karya ilmiah merupakan sebuah karya tulis yang membahas suatu permasalahan dengan didasari oleh hasil dari penelitian menggunakan metode dan sistematika penulisan yang baik dan isinya dapat dipertanggungjawabkan mengenai keilmiahan dan kebenarannya[1]. Dalam menulis karya ilmiah terdapat etika yang harus dijunjung tinggi oleh para peneliti yaitu menyebarluaskan atau mempublikasikan hasil dari karya ilmiah di media cetak atau elektronik[2],[3]. Hasil penelitian dapat dipublikasikan dalam jurnal ilmiah atau prosiding atau dalam bentuk buku, mencantumkan sumber dana penelitian jika bersedia, serta tidak melakukan pelanggaran seperti Falsifikasi, Fabrikasi, dan Plagiarisme[4],[5]. Di Indonesia banyak terjadi pelanggaran etika dalam penulisan ilmiah salah satunya adalah plagiarisme. Plagiarisme adalah tindakan secara sengaja untuk memperoleh nilai dari suatu karya ilmiah, dengan mengambil sebagian atau seluruh karya ilmiah pihak lain, tanpa menyatakan sumber secara tepat.

Plagiarisme dalam menulis karya ilmiah harus dihindari karena plagiarisme dapat merugikan banyak pihak, baik dari pihak korban plagiarisme maupun pihak yang melakukan plagiarisme karena terkena jeratan hukum yang belaku, serta pihak lembaga atau perguruan tinggi yang akan berdampak buruk bagi mutu perguruan tinggi. Peraturan tentang plagiarisme di Indonesia telah diatur dalam Permendiknas No.17 tahun 2010 tentang sanksi, pencegahan dan penanggulangan plagiat di perguruan tinggi. Karena semakin maraknya plagiarisme dalam penulisan karya ilmiah di Indonesia, dibutuhkan sebuah teknologi yang dapat mengetahui tingkat plagiarisme secara akurat. Adapun salah satu cara yang dapat digunakan untuk mengetahui presentase plagiarisme yaitu menggunakan text mining.

Text mining adalah sebuah teknik klasifikasi, dan text mining adalah sebuah variasi dari data mining yang berusaha mengelola data tekstual yang berjumlah besar untuk mendapatkan pola yang menarik di dalamnya[6],[7],[8]. Sedangkan menurut Karyadi, Gupta dan Lehal menyatakan bahwa text mining memiliki kemiripan dengan data mining, dimana data mining digunakan untuk mengelola data yang terstruktur yang terdapat didalam database, sedangkan text mining dapat digunakan untuk mengelola data semi terstruktur atau tidak terstruktur sama sekali seperti email, dokumen dan lain-

lain[9],[10]. Sehingga text mining adalah sebuah kegiatan mengekstrak informasi dari sumber tertulis untuk mendapatkan penemuan baru dari informasi yang belum diketahui didalamnya.

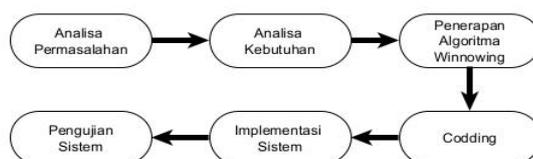
Algoritma winnowing adalah algoritma yang digunakan untuk mendeteksi kesamaan teks menggunakan fungsi hashing[11],[12]. Selain itu algoritma ini juga digunakan untuk mendeteksi plagiarisme dengan melakukan pengecekan kesamaan kata[13],[14]. Fakultas Ilmu Komputer UIGM Palembang telah mencetak banyak lulusan setiap tahunnya, dari jumlah mahasiswa fakultas ilmu komputer, setiap tahunnya rata-rata mahasiswa yang mengusulkan judul penelitian lebih dari 100 judul, mahasiswa diwajibkan untuk mengerjakan tugas akhir berupa sebuah karya ilmiah berbentuk skripsi sebagai syarat untuk menyelesaikan studi. Dari data penelitian 3 tahun terakhir yang berjumlah 325 judul, masih terdapat hasil penelitian yang dikerjakan memiliki plagiarisme atau kesamaan dengan judul penelitian yang sudah ada. Dari permasalahan tersebut, fakultas ilmu komputer terutama pada program studi membutuhkan suatu aplikasi sebagai alat bantu dalam pemecahan masalah, oleh karena itu dibangun suatu sistem yang dapat mengetahui similarity judul dengan menggunakan text minning dengan algoritma winnowing.

2. METODE PENELITIAN

Metode penelitian ini, pengumpulan data diperoleh dari data yang dikumpulkan data judul penelitian 3 tahun terakhir sebanyak 325 judul penelitian. Data tersebut diperoleh dari UPT Perpustakaan UIGM Palembang. Adapun metode penelitian dilakukan dengan metode atau algoritma winnowing.

2.1 Tahapan Penelitian

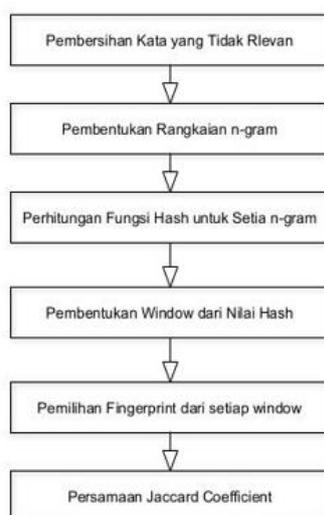
Pada gambar 1 merupakan tahapan penelitian yang dilakukan tentang penerapan text mining untuk identifikasi similarity pada judul penelitian menggunakan algoritma Winnowing[15]



Gambar 1. Tahapan Penelitian

2.2 Algoritma Winnowing

Algoritma winnowing merupakan teknik logika yang digunakan untuk mengidentifikasi plagiaris dengan teknik hashing. Masukan data dari algoritma winnowing berupa teks, dan akan mendapatkan luaran berupa nilai hash dalam bentuk perhitungan ASCII di setiap karakter[16]. Nilai hash untuk mendeteksi adanya plagiaris[17],[18]. Gambar 2 merupakan konsep algoritma winnowing.



Gambar 2. Konsep Algoritma Winnowing[19]

1. Pembersihan kata yang tidak relevan
dokumen harus melalui langkah pembersihan. Yaitu proses yang digunakan untuk menghapus karakter dan kata yang tidak diperlukan.
2. Pembentukan rangkaian n-gram
N-grams adalah susunan token dengan panjang n yang telah ditentukan. token berupa kata-kata, maupun karakter atau himpunan. Metode ini digunakan untuk memotong karakter huruf sejumlah n dari kata yang telah dibersihkan melalui preprocessing secara berkelanjutan dari awal hingga akhir.
3. Perhitungan fungsi hash pada n-gram
perhitungan ini untuk mengetahui nilai hash dari setiap gram dengan menggunakan rolling hash. Rolling Hash berfungsi sebagai penanda sebuah string dengan cara mentransformasi string tersebut menjadi nilai yang unik dengan fixed-length. Nilai yang dihasilkan disebut nilai hash dan fungsi untuk mengasilkan nilai tersebut adalah fungsi hash. Fungsi hash $H(c1..ck)$ didefinisikan sebagai berikut :

$$H(ck) = C1 * b(k - 1) + c2 * b(k - 2) + \dots + ck * b(k - k)$$

4. Pembentukan window dari nilai hash

Window berfungsi untuk memperoleh fingerprint yang digunakan pada pencocokan plagiasi melalui pembentukan substring dari nilai hash yang telah ditentukan.

5. Pemilihan fingerprint dari setiap window

Pemilihan berdasarkan nilai terkecil dari nilai hash tiap window. Jika terdapat dua nilai terkecil yang sama maka akan diambil nilai yang paling kanan.

6. Persamaan jaccard coefficient

Untuk mengukur presentase kemiripan teks dibutuhkan persamaan Jaccard Coefficient. Persamaan ini digunakan untuk menghitung kemiripan (similarity) dari kumpulan kata-kata yang telah dihitung nilai hash nya. Berikut ini rumus persamaan Jaccard Coefficient [20]:

$$\text{similarity} = \frac{\text{jumlah_fingerprint_sama}}{\text{Total_seluruh_fingerprint}} \times 100$$

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma Winnowing

Pada analisa perhitungan menggunakan algoritma winnowing digunakan 2 judul penelitian yang digunakan sebagai perbandingan untuk mengetahui persentase kesamaan judul penelitian menggunakan text mining dengan algoritma winnowing seperti pada tabel 1 sebagai berikut

a. Perhitungan Text Preprocessing

Perhitungan ini bertujuan untuk melakukan penghapusan karakter yang tidak diperlukan dan mengubah menjadi huruf kecil. Perhitungan text preprocessing seperti pada tabel 1 sebagai berikut :

Tabel 1. Perhitungan Text Preprocessing

Judul 1	Judul 2
pembuatan sistem informasi eksekutif kinerja dosen di perguruan tinggi fakultas ekonomi universitas musi rawas dengan menggunakan teknologi berbasis web	sistem informasi eksekutif kinerja dosen fakultas ekonomi universitas musi rawas berbasis web

b. Perhitungan Feature Selection

Pada perhitungan ini terdapat 2 tahapan yaitu stemming dan stopwords removal. Stemming menggunakan algoritma Nazief & Adriani berguna untuk mengubah kata menjadi kata baku seperti kata “penerapan” menjadi “terap”. Perhitungan Stemming dapat dilihat pada tabel 2.

Tabel 2. Perhitungan Stemming

Judul 1	Judul 2
buat sistem informasi eksekutif kinerja dosen di guru tinggi fakultas ekonomi universitas mus rawas dengan guna teknologi basis web	sistem informasi eksekutif kinerja dosen fakultas ekonomi universitas mus rawas basis web

Stop word Removal yaitu penghapusan kata yang bukan kata inti dari sebuah kalimat seperti menghapus kata “dan”, “atau”, “untuk” dan “jika”. Perhitungan stop word removal dapat dilihat pada tabel 3.

Tabel 3. Perhitungan Stop Word Removal

Judul 1	Judul 2
buat sistem informasi eksekutif kinerja dosen guru tinggi fakultas ekonomi universitas mus rawas guna teknologi basis web	sistem informasi eksekutif kinerja dosen fakultas ekonomi universitas mus rawas basis web

c. Perhitungan n-gram

Perhitungan n-gram adalah memotong kata sepanjang n yang telah di tentukan. Pada perhitungan ini menggunakan n-gram = 5. Hasil perhitungan n-gram pada tabel 4 sebagai berikut.

Tabel 4. Perhitungan N-Gram

Judul 1	Judul 2
buats uatsi atsis tsist siste istem stemi temin eminf minfo infor nform forma ormas rmasi masie asiek sieks iekse eksek kseku sekut ekuti kutif utifk tifki ifkin fkine kiner inerj nerja erjad rjado jados adose dosen oseng sengu engur nguru gurut uruti rutin uting tingg inggi nggif ggifa gifak ifaku fakul akult kulta	siste istem stemi temin eminf minfo infor nform forma ormas rmasi masie asiek sieks iekse eksek kseku sekut ekuti kutif utifk tifki ifkin fkine kiner inerj nerja erjad rjado jados adose dosen osenf senfa enfak nfaku fakul akult kulta ultas ltase

ultas ltase tasek aseko sekon ekono konom onomi nomiu omiun miuni iuniv unive niver ivers versi ersit rsita sitas itasm tasmu asmus smusr musra usraw srawa rawas awasg wasgu asgun sguna gunat unate natek atekn tekno eknol knolo nolog ologi logib ogiba gibas ibasi basis asisw siswe isweb	tasek aseko sekon ekono konom onomi nomiu omiun miuni iuniv unive niver ivers versi ersit rsita sitas itasm tasmu asmus smusr musra usraw srawa rawas awasb wasba asbas sbasi basis asisw siswe isweb
--	--

d. Perhitungan Hash untuk Setiap n-gram

Menghitung hash dengan menggunakan Rolling Hash untuk masing-masing n-gram. Pada n-gram judul pertama "buats" dengan basis bilangan prima (b) = 2, dengan rangkaian panjang n-gram = 5 adalah sebagai berikut.

$$H(ck) = c1 * b(k - 1) + c2 * b(k - 2) + \dots + ck * b(k - k)$$

$$H(\text{buats}) = \text{asci}(\text{b}) * 2^4 + \text{asci}(\text{u}) * 2^3 + \text{asci}(\text{a}) * 2^2 + \text{asci}(\text{t}) * 2^1 + \text{asci}(\text{s}) * 2^0$$

$$= 98 * 2^4 + 117 * 2^3 + 97 * 2^2 + 116 * 2^1 + 115 * 2^0$$

$$= 98 * 16 + 117 * 8 + 97 * 4 + 116 * 2 + 115 * 1$$

$$= 1568 + 936 + 388 + 232 + 115$$

$$= 3239$$

$$H(\text{uatsi}) = \text{asci}(\text{u}) * 2^4 + \text{asci}(\text{a}) * 2^3 + \text{asci}(\text{t}) * 2^2 + \text{asci}(\text{s}) * 2^1 + \text{asci}(\text{i}) * 2^0$$

$$= 117 * 2^4 + 97 * 2^3 + 116 * 2^2 + 115 * 2^1 + 105 * 2^0$$

$$= 117 * 16 + 97 * 8 + 116 * 4 + 115 * 2 + 105 * 1$$

$$= 1872 + 776 + 464 + 230 + 105$$

$$= 3447$$

$$H(\text{atsis}) = \text{asci}(\text{a}) * 2^4 + \text{asci}(\text{t}) * 2^3 + \text{asci}(\text{s}) * 2^2 + \text{asci}(\text{i}) * 2^1 + \text{asci}(\text{s}) * 2^0$$

$$= 97 * 2^4 + 116 * 2^3 + 115 * 2^2 + 105 * 2^1 + 115 * 2^0$$

$$= 97 * 16 + 116 * 8 + 115 * 4 + 105 * 2 + 115 * 1$$

$$= 1552 + 928 + 460 + 210 + 115$$

$$= 3265$$

$$H(\text{tsist}) = \text{asci}(\text{t}) * 2^4 + \text{asci}(\text{s}) * 2^3 + \text{asci}(\text{i}) * 2^2 + \text{asci}(\text{s}) * 2^1 + \text{asci}(\text{t}) * 2^0$$

$$= 116 * 2^4 + 115 * 2^3 + 105 * 2^2 + 115 * 2^1 + 116 * 2^0$$

$$= 116 * 16 + 115 * 8 + 105 * 4 + 115 * 2 + 116 * 1$$

$$= 1856 + 920 + 420 + 230 + 116$$

$$= 3542$$

$$H(\text{siste}) = \text{asci}(s)*24 + \text{asci}(i)*23 + \text{asci}(s)*22 + \text{asci}(t)*21 + \text{asci}(e)*20$$

$$= 115*24 + 105*23 + 115*22 + 116*21 + 101*20$$

$$= 115*16 + 105*8 + 115*4 + 116*2 + 101*1$$

$$= 1840 + 840 + 460 + 232 + 101$$

$$= 3473$$

Adapun hasil perhitungan dari semua hash setiap N-gram adalah seperti pada tabel 5.

Tabel 5. Perhitungan Hash untuk Setiap N-Gram

Judul 1	Judul 2
3239 3447 3265 3542 3473 3375 3495	3473 3375 3495 3420 3230 3339
3420 3230 3339 3304 3357 3291 3433	3304 3357 3291 3433 3419 3291
3419 3291 3201 3413 3247 3241 3367	3201 3413 3247 3241 3367 3426
3426 3277 3424 3531 3423 3244 3229	3277 3424 3531 3423 3244 3229
3308 3298 3333 3246 3371 3209 3127	3308 3298 3333 3246 3371 3209
3260 3423 3411 3256 3397 3390 3589	3127 3260 3422 3389 3205 3295
3544 3543 3445 3283 3308 3193 3197	3178 3208 3409 3509 3375 3401
3215 3178 3208 3409 3509 3375 3401	3201 3408 3247 3371 3423 3411
3201 3408 3247 3371 3423 3411 3412	3412 3377 3384 3509 3388 3371
3377 3384 3509 3388 3371 3487 3314	3487 3314 3493 3453 3335 3427
3493 3453 3335 3427 3257 3524 3465	3257 3524 3465 3561 3475 3385
3561 3475 3385 3225 3463 3228 3449	3220 3433 3173 3347 3129 3241
3334 3473 3309 3208 3423 3242 3363	3479 3376
3405 3395 3336 3313 3189 3187 3129	
3241 3479 3376	

e. Perhitungan Window

Perhitungan window digunakan untuk memotong hasil hash n-gram sepanjang window yang telah ditentukan. perhitungan ini menggunakan window = 5

f. Pemilihan Fingerprint untuk Window

Pemilihan angka terkecil dari masing-masing window. Untuk seterusnya hasil pemilihan angka terkecil ini dapat diistilahkan sebagai fingerprint. Hasil dari

fingerprint dari judul 1 terdapat 68 window, dan dari judul 2 terdapat 78 window.

Fingerprint dari judul 1 antara lain :

Tabel 6. Pemilihan Fingerprint dari Setiap Window

Judul 1	Judul 2
3239 3265 3265 3375	3230 3230 3230
3230 3230 3230	3230 3230 3291
3230 3230 3291 3291	3291 3291 3201
3291 3201 3201	3201 3201 3201
3201 3201 3201 3241	3201 3241 3241
3241 3241 3277	3241 3277 3277
3277 3244 3229 3229	3244 3229 3229
3229 3229 3229	3229 3229 3229
3246 3209 3127 3127	3246 3209 3127
3127 3127 3127	3127 3127 3127
3256 3256 3256 3256	3127 3205 3178
3390 3390 3283	3178 3178 3178
3283 3193 3193 3193	3178 3208 3201
3178 3178 3178	3201 3201 3201
3178 3178 3208 3201	3201 3247 3247
3201 3201 3201	3371 3377 3377
3201 3247 3247 3371	3377 3371 3371
3377 3377 3377	3314 3314 3314
3371 3371 3314 3314	3314 3314 3257
3314 3314 3314	3257 3257 3257
3257 3257 3257 3257	3257 3385 3220
3257 3385 3225	3220 3173 3173
3225 3225 3225 3225	3129 3129 3129
3228 3228 3208	3129
3208 3208 3208 3208	
3242 3242 3313	
3189 3187 3129 3129	
3129 3129	

g. Persamaan Jaccard Coefficient

untuk dapat menentukan koefisien ini diperlukan perhitungan dari fingerprint yang sudah didapat antara lain:

jumlah window pada judul 1 = 97
 jumlah window pada judul 2 = 70
 union (gabungan) fingerprint 1 dan 2 = 167
 intersection (fingerprint yang sama) = 70
 (union - intersection) = 97

Dengan menggunakan data diatas dapat dihitung persentase kesamaan judul dengan menggunakan jaccard coefficient yaitu:

Jaccard coeficiend = $(\text{intersection} / (\text{union} - \text{intersection})) * 100\%$
 Jaccard coeficiend = $70/97 * 100\%$
 Hasil presentase kesamaan = 72.16 %

Adapun hasil yang didapat dengan menggunakan kombinasi n-gram dan window yang berbeda seperti pada tabel 7 sebagai berikut:

Tabel 7. Daftar Hasil Perhitungan

Jumlah N-Gram	Jumlah window	Presentase kemiripan
3	3	88.17%
3	4	90.11%
3	5	87.91%
3	6	89.98%
3	7	87.64%
4	3	94.38%
4	4	94.32%
4	5	94.25%
4	6	85.56%
4	7	87.5%
5	3	69.31%
5	4	70.71%
5	5	72.16%
5	6	66.67%
5	7	69.79%
6	3	65.69%
6	4	68.69%
6	5	70.1%
6	6	66.33%
6	7	69.47%
7	3	60.58%
7	4	61.76%

Jumlah N-Gram	Jumlah window	Presentase kemiripan
7	5	61.39%

3.2 Perhitungan Nilai N-Gram dan Window

Perhitungan ini untuk mengetahui kombinasi antara n-gram dan window mana yang memiliki presentase kesamaan yang paling tepat dengan melakukan perhitungan akurasi kepada 2 judul yang telah melewati tahap pembersihan. Kombinasi n-gram dan window yang diuji dengan range dimulai dari angka 3 sampai dengan 7. Penggunaan n-gram dan window 1 dan 2 tidak dianjurkan karena memiliki tingkat error yang sangat tinggi. Apabila judul yang diajukan memiliki karakter yang lebih sedikit dibandingkan judul yang sudah ada menyebabkan hasil perhitungan mencapai diatas ratusan persen. Penggunaan n-gram dan window diatas 7 juga tidak dianjurkan karena hasil yang didapat sangat kecil dan tidak dapat dijadikan patokan untuk menghitung presentase kesamaan judul. mengetahui n-gram dan window yang sesuai.

Tabel 8. Hasil text Preprocessing

Judul pertama	Judul kedua
Pebuatan Sistem Informasi Eksekutif Kinerja Dosen Di Perguruan Tinggi Fakultas Ekonomi Universitas Musi Rawas Dengan Menggunakan Teknologi Berbasis Web	Sistem Informasi Eksekutif Kinerja Dosen Fakultas Ekonomi Universitas Musi Rawas Berbasis Web
Text preprocessing Judul pertama	Text preprocessing Judul kedua
buat sistem informasi eksekutif kinerja dosen guru tinggi fakultas ekonomi universitas mus rawas guna teknologi basis web	sistem informasi eksekutif kinerja dosen fakultas ekonomi universitas mus rawas basis web

Hasil presentase kesamaan judul penelitian yang di dapat adalah:

$$\text{Akurasi} = \text{jumlah judul 2} / \text{jumlah judul 1} * 100\%$$

$$\text{Akurasi} = 105/78 * 100\%$$

$$\text{Akurasi} = 74,28\%$$

Dari perhitungan di atas, dapat diketahui selisih antara perhitungan akurasi dengan perhitungan winnowing yang dapat dilihat pada tabel 9.

Tabel 9. Pengujian N-Gram dan Window

Jumlah n-gram	Jumlah window	Presentase kemiripan	Selisih
3	3	88,17%	13,89%
3	4	90,11%	15,83%
3	5	87,91%	13,63%
3	6	89,98%	15,7%
3	7	87,64%	13,36%
4	3	94,38%	20,1%
4	4	94,32%	20,04%
4	5	94,25%	19,97%
4	6	85,56%	11,28%
4	7	87,5%	13,22%
5	3	69,31%	4,97%
5	4	70,71%	3,57%
5	5	72,16%	2,12%
5	6	66,67%	7,61%
5	7	69,79%	4,49%
6	3	65,69%	8,59%
6	4	68,69%	5,59%
6	5	70,1%	4,18%
6	6	66,33%	7,95%
6	7	69,47%	4,81%
7	3	60,58%	13,7%
7	4	61,76%	12,52%
7	5	61,39%	12,89%
7	6	56,31%	17,97%
7	7	57,43%	16,85%

Dari perhitungan akurasi di atas didapatkan persentase kesamaan yang mendekati angka 74,28% yaitu kombinasi n-gram = 5 dan window = 5. Kombinasi ini menghasilkan persentase kesamaan 72,16% dengan selisih 2,12%. Dari pengujian diatas dapat ditarik kesimpulan bahwa kombinasi n-gram dan winndow yang tepat untuk digunakan dalam penelitian ini adalah n-gram = 5 dan window = 5.

3.3 Implementasi Sistem

a. Proses Input Judul Penelitian

Halaman ini merupakan menu yang digunakan untuk melakukan input judul penelitian yang ingin diajukan. Adapun inputan yang harus dimasukkan di halaman ini adalah judul penelitian yang akan diperiksa persentase kesamaannya.

Gambar 3. Proses Input Judul Penelitian

b. Proses Ubah Status Pemeriksaan Judul Penelitian

Pada halaman ini merupakan proses perubahan status pemeriksaan judul skripsi jika judul skripsi telah dilakukan pemeriksaan terhadap kesamaan judul skripsi yang ada pada basis data.

Gambar 4. Halaman Ubah Status Pemeriksaan Judul Penelitian

c. Hasil Perhitungan

Halaman ini digunakan untuk menampilkan hasil presentase kesamaan judul menggunakan algoritma winnowing dengan data yang sudah melalui tahap pembersihan dengan text preprocessing dan feature selection. Pada halaman ini, hasil perhitungan hanya menampilkan 5 judul teratas yang memiliki presentase kesamaan tertinggi terhadap judul yang diajukan. Adapun dihalaman ini terdapat tombol detail yang berfungsi untuk menampilkan halaman detail perhitungan, dan tombol cetak untuk mencetak laporan hasil perhitungan.

No	Judul	Persentase	Aksi
1	SISTEM INFORMASI EKSEKUTIF KINERJA DOSEN FAKULTAS EKONOMI UNIVERSITAS MUSI RAWAS BERBASIS WEB	72,16%	Detail
2	PERANCANGAN SISTEM INFORMASI MANAJEMEN PROSES BIMBINGAN SKRIPSI PADA FAKULTAS EKONOMI UNIVERSITAS MUSI RAWAS	46,65%	Detail
3	SISTEM INFORMASI EKSEKUTIF BIDANG AKADEMIK PADA PROGRAM STUDI ILMU SOSIAL DAN POLITIK UNIVERSITAS MUSI RAWAS	26,81%	Detail
4	RANCANG BANGUN SISTEM INFORMASI MARKETPLACE UNTUK JASA WEDDING ORGANIZER BERBASIS WEB	18,18%	Detail
5	SISTEM INFORMASI MANAJEMEN PENGAEDIAN KEPADA MASYARAKAT BERBASIS WEB RESPONSIVE PADA UNIVERSITAS MUSI RAWAS	16,06%	Detail

Gambar 5. Halaman Hasil Perhitungan

3.4 Pengujian Sistem

Pada tahapan ini dilakukan pengujian terhadap sistem yang telah dibangun. Dimana teknik pengujian ini menggunakan teknik user acceptance testing (UAT) dimana pengujian ini dilakukan untuk mengetahui apakah sistem yang dibangun telah sesuai dengan kebutuhan pengguna. Adapun hasil pengujian sebagai berikut:

- Analisis pertanyaan pertama adalah $5/1=5$ Nilainya $5/5 \times 100=100\%$
- Analisis pertanyaan kedua adalah $4/1=4$ Nilainya $4/5 \times 100=80\%$
- Analisis pertanyaan ketiga adalah $4/1=4$ Nilainya $4/5 \times 100=80\%$
- Analisis pertanyaan keempat adalah $4/1=4$ Nilainya $4/5 \times 100=80\%$
- Analisis pertanyaan kelima adalah $5/1=5$ Nilainya $5/5 \times 100=100\%$

Adapun kesimpulan dari pengujian dengan menggunakan teknik UAT adalah implementasi text mining dengan menggunakan algoritma Winnowing ini memiliki proses yang sangat sesuai, tampilan yang sesuai, menu- menu yang mudah dipahami, dan sangat sesuai dengan kebutuhan pengguna.

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan maka diperoleh suatu kesimpulan pada penelitian ini yaitu untuk dapat mengetahui presentase kesamaan judul diperlukan text mining dengan algoritma winnowing. Hasil presentase kesamaan judul penelitian menggunakan algoritma winnowing ini sangat dipengaruhi dari nilai n-gram dan window. persentase kesamaan yang mendekati angka 74,28% yaitu kombinasi n-gram = 5 dan window = 5. Kombinasi ini menghasilkan persentase kesamaan 72,16% dengan selisih 2,12%. Pada hasil penelitian ini juga dapat melakukan pengecekan plagiarisme atau deteksi duplikasi dalam basis data dari berbagai bahasa dan format teks.

REFERENCES

- [1] Nirwana and Abd. Rahim Ruspa, "Kemampuan Menulis Karya Tulis Ilmiah Mahasiswa Prodi Informatika Universitas Cokroaminoto Palopo," *J. Onoma Pendidikan, Bahasa, dan Sastra*, vol. 6, no. 1, pp. 557–566, 2020, doi: 10.30605/onoma.v6i1.277.
- [2] J. Z. Hafizd, "Implementasi Peran Mahasiswa Sebagai Agent Of Change Melalui Karya Tulis Ilmiah," *Dimasejati J. Pengabd. Kpd. Masy.*, vol. 4, no. 2, pp. 175–184, 2022.
- [3] R. C. Hastari, "Pelatihan Penelusuran Referensi dan Pencegahan Plagiasi Pada Kegiatan Diklat Dasar Karya Tulis Ilmiah," *Cendekia J. Pengabd. Masy.*, vol. 3, no. 1, p. 18, 2021, doi: 10.32503/cendekia.v3i1.1582.
- [4] R. K. Wibowo *et al.*, "Penerapan Algoritma Winnowing Untuk," *Techno.Com*, vol. 15, no. 4, p. 304, 2016.
- [5] Y. Rohmawati, N. Hidayah, and Y. Hotifah, "Peningkatan Mutu Profesi Guru Melalui Program Menulis Sebgaai Tanggungjawab Etika Keilmuan," *Pros. Semin. Nas.*, pp. 36–43, 2023.
- [6] N. L. Ratniasih, M. Sudarma, and N. Gunantara, "Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 3, p. 13, 2017, doi: 10.24843/mite.2017.v16i03p03.
- [7] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [8] A. S. Aribowo and S. Khomsah, "Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19) Implementasi Text Mining Untuk Deteksi Emosi Menggunakan Metode Leksikon (Studi Kasus: Twit Tentang Covid-19)," *J. Inform. dan Teknol. Inf.*, vol. 18, no. 1, pp. 49–60, 2021, doi: 10.31515/telematika.v18i1.4341.
- [9] S. Karyadi, H. Yasin, and M. A. Mukid, "Analisis Kecenderungan Informasi Dengan Menggunakan Metode Text Mining," *J. Gaussian*, vol. 5, no. 4, pp. 763–770, 2016, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [10] N. Putra Bayu Pratama, Mustaqiem, and Minarni, "Pengembangan Sistem Informasi Pengelolaan Judul Skripsi dan Tugas Akhir dengan Fitur Deteksi Kemiripan Menggunakan Algoritma Winnowing," *Terap. Inform. Nusant.*, vol. 2, no. 5, pp. 271–278, 2021.
- [11] J. Jarwati, A. C. Prihandoko, and W. E. Yulia R, "Penerapan Algoritma Winnowing Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)," *Berk. Sainstek*, vol. 5, no. 1, p. 11, 2017, doi: 10.19184/bst.v5i1.5369.
- [12] N. ALAMSYAH, "Perbandingan Algoritma Winnowing Dengan Algoritma Rabin Karp Untuk Mendeteksi Plagiarisme Pada Kemiripan Teks Judul Skripsi," *Technol. J. Ilm.*, vol. 8, no. 3, p. 124, 2017, doi: 10.31602/tji.v8i3.1116.
- [13] A. RK, Wibowo., K, "Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa," *Techno.Com*, vol. 15, no. 4, pp. 303–311, 2016.
- [14] W. Hidayat, E. Utami, and A. D. Hartanto, "Pemilihan Parameter Terbaik pada Algoritma Winnowing dalam Mendeteksi Tingkat Kesamaan Dokumen Bahasa Indonesia," *Creat. Inf. Technol. J.*, vol. 7, no. 2, p. 119, 2021, doi: 10.24076/citec.2020v7i2.256.
- [15] I. P. Handayani *et al.*, "SIMPLE ADDITIVE WEIGHTING PADA SISTEM KELAYAKAN," vol. 7, pp. 267–278, 2024.
- [16] I. Widaningrum, D. Mustikasari, R. Arifin, and H. A. Pratiwi, "Evaluation of the accuracy of winnowing, rabin karp and knuth morris pratt algorithms in plagiarism detection applications," *J. Phys. Conf. Ser.*, vol. 1517, no. 1, 2020, doi: 10.1088/1742-6596/1517/1/012093.
- [17] M. Peka Rih, A. A. Pekuwali, and D. A. Sitaniapessy, "Algoritma Winnowing untuk Mendeteksi Kesamaan Judul Skripsi Teknik Informatika," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 42–52, 2022, doi: 10.57152/malcom.v2i2.436.
- [18] M. S. Ramli, S. Cokrowibowo, and M. F. Rustan, "Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing," *J. Appl. Comput. Sci. Technol.*, vol. 2, no. 2, pp. 108–112, 2021, doi: 10.52158/jacost.v2i2.177.
- [19] S. Sunardi, A. Yudhana, and I. A. Mukaromah, "Implementasi Deteksi Plagiarisme Menggunakan Metode N-Gram Dan Jaccard Similarity Terhadap Algoritma Winnowing," *Transmisi*, vol. 20, no. 3, p. 105, 2018, doi: 10.14710/transmisi.20.3.105-110.
- [20] K. C. Kirana, J. A. Prasetya, and S. Wibawanto, "Analisis ChatGPT dan Perplexity pada Mata Pelajaran Dasar Program Keahlian dengan Jaccard Similarity," *Belantika Pendidik.*, vol. 6, no. 1, pp. 36–46, 2023.