

Deteksi Plagiarisme Abstrak Skripsi dengan Menggunakan Algoritma Rabin Karp (Studi Kasus: Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang)

Indra Gunawan
Universitas Singaperbangsa Karawang
Karawang, Indonesia
indra.16114@student.unsika.ac.id

Ultach Enri
Universitas Singaperbangsa Karawang
Karawang, Indonesia
ultach.enri@staff.unsika.ac.id

Yuyun Umaidah
Universitas Singaperbangsa Karawang
Karawang, Indonesia
yuyun.umaidah@staff.unska.ac.id

Abstract— Pada lingkup pendidikan khususnya perguruan tinggi plagiarisme sering terlihat. Umumnya plagiarisme terjadi karena rasa malas dan ingin cepat dalam menyelesaikan urusan tugasnya. Algoritma Rabin Karp merupakan algoritma pencarian *string*, algoritma ini digunakan untuk mendeteksi plagiarisme pada teks. Tujuan penelitian mengetahui hasil evaluasi yang didapat dari proses Algoritma Rabin Karp. Data penelitian akan melewati semua tahapan *preprocessing* (*case folding*, *tokenizing*, *filtering*, dan *stemming*) dan melewati sebagian tahapan *preprocessing* (*case folding*), *k-gram* yang diuji yaitu 2gram, 3gram, 4gram, 5gram, dan 6gram kemudian melewati tahapan *hashing* dan mendapatkan nilai *fingerprint* kemudian diuji tingkat kemiripannya menggunakan *Dice Similarity Coefficient*. Metode penelitian yang digunakan yaitu metode *Text Mining* yang memiliki tahapan Akuisisi, *Text Preprocessing*, *Modeling*, dan Evaluasi. Dari data yang digunakan menghasilkan nilai rata-rata total kemiripan 86.84% pada 2gram, 69.56% pada 3gram, 56.06% pada 4gram, 48.71% pada 5gram, dan 44.30% pada 6gram. hasil dari tahapan *Preprocessing* dengan hasil dari tahapan Sebagian *Preprocessing*, memiliki perbedaan yaitu, hasil tahapan *Preprocessing* lebih kecil persentase kemiripannya daripada hasil sebagian *Preprocessing*, ini disebabkan penghilangan kata pada tahapan *filtering* dan perubahan kata pada tahapan *stemming*. Dapat disimpulkan bahwa dari data yang digunakan terlihat adanya tindakan plagiarisme pada abstrak, hal ini didukung dengan adanya data yang memiliki nilai kemiripan hingga 100%.

Kata kunci — Plagiarisme, Rabin Karp, Text Mining

I. PENDAHULUAN

Pada lingkup pendidikan khususnya pada perguruan tinggi plagiarisme masih sering kita lihat. Umumnya pelaku plagiarisme adalah mahasiswa, plagiarisme ini dilakukan saat penggerjaan tugas yang diberikan oleh dosen dan kemudian mahasiswa melakukan teknik copy & paste baik itu berasal dari internet dimana semua orang dapat mengakses maupun hasil karya orang lain atau hasil plagiarisme dari temannya, karena hal tersebut dapat memicu tindakan plagiarisme pada penulisan skripsi mahasiswa nantinya. Plagiarisme atau biasa disebut plagiatis adalah pencurian hasil karya orang lain dan menjadikannya milik sendiri [1]. Adapun klasifikasi berdasarkan proporsi atau persentase kata/kalimat/paragraf yang dibajak yaitu : (1) Plagiarisme ringan dimana kata yang dibajak < 30%, (2) plagiarisme sedang dimana kata yang dibajak berkisar dari 30% hingga 70%, dan (3) plagiarisme berat dimana kata yang dibajak yaitu melebihi dari 70% [2].

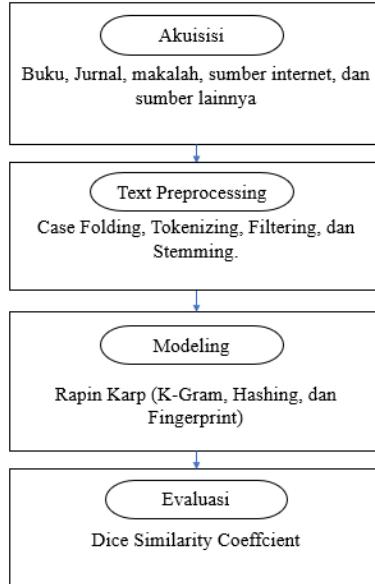
Penelitian yang dilakukan oleh Sibarani, Magdalena, dan Dharma yaitu Analisa Perbandingan Sistem Pendekripsi Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing Dan Algoritma Rabin Karp menyimpulkan bahwa berdasarkan pengujian sistem, tingkat akurasi untuk kemiripan judul dan beserta isinya dengan Algoritma winnowing dan Algoritma Rabin karp mendekati 75%, dan menghasilkan persentase plagiarisme sebesar 97,22% [3]. Algoritma Rabin Karp merupakan algoritma pencarian string yang menggunakan fungsi hashing untuk membandingkan string yang dicari (m) dengan substring pada teks (n) [4].

Identifikasi Bentuk Tindak Plagiatis Pada Penulisan Skripsi Mahasiswa Program Studi Pendidikan Fisika Unsyiah menyimpulkan bahwa penelitian tersebut meneliti dengan menggunakan 134 skripsi mahasiswa Program Studi Pendidikan Fisika tahun 2016, dan menggunakan bantuan software <http://www.smallseotools.com/> menyimpulkan bahwa dari 134 skripsi yang diteliti pada landasan teoritis teridentifikasi plagiat penuh sebesar 80,55% [5].

Langkah untuk mendeteksi plagiarisme yaitu menggunakan konsep kemiripan dokumen dengan menggunakan Text Mining. Text Mining itu sendiri adalah ilmu untuk mencari dan mengukur data yang berupa teks yang bertujuan untuk mencari kata-kata yang dapat merepresentasikan isi dan maksud dari teks tersebut kemudian melakukan penganalisaan lebih lanjut [1]. Penelitian ini menggunakan algoritma Rabin Karp dalam mendeteksi plagiarisme pada teks karena lebih efektif dan praktis dibandingkan dengan algoritma yang lain, algoritma ini melakukan prinsip perhitungan dengan menghitung nilai hash untuk mendapatkan hasil gram-nya kemudian proses ini akan dilakukan secara berkelanjutan [3]. K-gram merupakan rangkaian terms yang panjang dengan nilai panjang k. Metode K-gram ini mengambil setiap karakter huruf dengan jumlah nilai k dari sebuah teks secara terus menerus kesebelah kanan teks dari awal teks hingga akhir teks. K-gram akan memecah string ke dalam potongan substring sebanyak k [6]. Algoritma Rabin Karp ini adalah algoritma yang dapat mencari kata dengan banyak pola sekaligus [7].

II. METODOLOGI

Metodologi yang digunakan pada penelitian ini yaitu menggunakan metode Text Mining. Proses pada metode Text Mining melalui empat tahapan yaitu akuisisi, text preprocessing, modeling, dan evaluasi.



Gambar 1 Metode Text Mining

Gambar 1 memiliki tahapan yaitu peneliti akan melakukan akuisisi yang didalamnya akan melakukan pencarian referensi, kemudian akan melakukan text preprocessing atau mentransformasikan kata. Kemudian akan malakukan Modeling atau pengimplementasian algoritma Rabin Karp tahapannya yaitu k-gram atau pengambilan setiap karakter huruf dengan jumlah nilai k dari sebuah teks secara terus menerus kesebelah kanan teks dari awal teks hingga akhir teks, Hashing merupakan cara untuk mengubah karakter string menjadi integer yang disebut nilai hash, Fingerprint merupakan bagian dari nilai hash dari proses hashing yang memiliki nilai yang sama, dan terakhir akan mendapatkan suatu kesimpulan menggunakan Dice Similarity Coeffcient. Dice Similarity Coeffcient adalah metode untuk menghitung kemiripan fingerprint dengan cara pendekatan K-Gram yang digunakan pada kedua teks yang diuji [8].

III. HASIL DAN PEMBAHASAN

Data abstrak skripsi yang diperoleh dari tahun 2018-2019, Data abstrak yang diperoleh yaitu sekitar 470 data abstrak skripsi. berikut ini merupakan data abstrak yang digunakan yaitu.

A	B	C	D	E	F	G	H	I	J	K
ABU BAKAR	Prototype, CRM, PHP, Codeigniter, MySQL	2018 DESAIN DAN IMPLEMENTASI CUSTOMER RELATIONS	Pengguna internet di Indonesia tiap tahun semakin bertambah, dari							
ALIH GUNAWAN	Rekam Medis, Tujuan Rekam Medis, Fungsi R	2018 RANCANG BANGUN APLIKASI REKAM MEDIS PASIEN	Rekam medis merupakan sebuah berkas yang didalamnya berisikan c							
ARIES YUDHA PUTRA PI	Obesitas, Sistem Pakar, Forward Chaining, Al	2018 SISTEM PAKAR PENGETAHUAN OBESITAS	Obesitas di indonesia cukup besar, banyak orang yang mem							
ABDUL ROHMAN	Hasil Biografi, spread spectrum, Huffman code	2018 IMPLEMENTASI METODE SPREAD SPECTRUM PADA PERKENALAN KINERJA GURU MENGGUNAKAN METOD	Hasil biografi merupakan hasil yang diperoleh							
ABDUL ROHMAN	Hasil Biografi, spread spectrum, Huffman code	2018 IMPLEMENTASI METODE SPREAD SPECTRUM PADA PERKENALAN KINERJA GURU MENGGUNAKAN METOD	Hasil biografi merupakan hasil yang diperoleh							
ABDULLAH LUBIS	Penilaian Kinerja, Analytical Hierarchy Process	2018 PENILAIAN KINERJA GURU MENGGUNAKAN METOD	Peran guru sebagai pengajar dalam proses belajar mengajar sangatlah							
ADE ALFIANSYAH	Augmented Reality, Media Pembelajaran, Or	2018 IMPLEMENTASI TEKNOLOGI AUGMENTED REALITY	Perkenalan ilmu pengetahuan dan teknologi semakin berkembang							
ADE NURHAYATI	Data Mining, C4.5, Padi, Kegagalan Panen,	2018 MIN-MAX NORMALIZATION UNTUK PREDIKSI TINGKARAN	terkenal sebagai kota hambang padi, karena menang hasil							
ADITIYA NUGROHO	Aplikasi pemasaran berbasis mobile Andro	2018 PENGEMBANGAN SISTEM ORDER BARANG BERBASIS PENELITI	Perum Peruri, Levenshtein Distance, Monitor							
ADJI WICAKSONO	Perum Peruri, Levenshtein Distance, Monitor	2018 PENERAPAN FITTING BENCARIAN DESEN ALGORITMA PENELITI	Peneliti sistem bekerja 24 jam non-stop untuk mengatasi							
EFENDI AFIFAH	metrik SED, metrik SED, metrik SED	2018 PENERAPAN METODE WEIGHTED PRODUCT (WNP) UNTUK PENGETAHUAN	metrik SED, metrik SED, metrik SED							
AGUNG FEIDYANTO	Penerapan Sistem Pendukung Keputusan, Se	2018 PENERAPAN METODE WEIGHTED PRODUCT (WNP) UNTUK PENGETAHUAN	metrik SED, metrik SED, metrik SED							
AGUNG MAULANA SALI	e-commerce, customer relationship manager	2018 APLIKASI E-COMMERCE DENGAN BENERAPAN CUST SOARING	dengan perkembangan teknologi yang semakin canggih, saat							
AGUNG RIYANTO	Android, BRI, Brilink, Google Map API, Locati	2018 APLIKASI INFOMASI LOKASI AGEN BRILINK MENG	sekarang ini banyak yang menggunakan teknologi							
AGUNG SIDDIO PERMAT	Perusahaan, Informasi, Maintenance, Produk	2018 PENERAPAN METODE MVC DALAM APLIKASI PENG	perkembangan teknologi yang semakin canggih, saat							
AGUS WIBAWA	Logika Fuzzy, Mandiri, Arduino, suhu, pH	2018 PENERAPAN FUNGSI FUZZY UNTUK PENGUKURAN	semakin meningkat							
AGUS ZAENUDIN	Algoritma naive bayes, Weka, percentage spl	2018 KLASIFIKASI PERPUTERAN PENDUDUK	KPenelitian ini merupakan studi yang sangat populer di							
AJAH ANGGONO	Knowledge Management System, Pariwisata,	2018 PENERAPAN METODE WEIGHTED PRODUCT (WNP) UNTUK PENGETAHUAN	Indonesia dan dunia internasional. Selain itu, sebagian besar							
Amelia Afifah	Analisa dan Interpretasi	2018 Penerapan Augmented Reality untuk pengetahuan	pengetahuan teknologi							
AJAT NURAJAT	Penilaian kinerja, SPK, Metode Simple Addit	2018 SISTEM PENDUKUNG KEPUTUSAN PENILIAHAN KINER	Augmented Reality, memperbaiki suatu kondisi antara							
AJAT ANGGONO	PHP, Codeigniter, AdminLTE, Facebook Adve	2018 EFektivitas IKLAN PADA FACEBOOK ADVERTISING Abstrak	dan antara							
LAKEHAD, ATI GHIFERON	Sistem,Pendukung Keputusan, Simile, Adriti,	2018 PENGERAPAN SIMPLE ADDITIVE WEIGHTING PADA PE	Facebook Advertising							

Gambar 2 Data Skripsi Fakultas Ilmu Komputer

Data yang digunakan akan diambil kata kunci yang telah diubah menjadi kode yaitu ak (Android-Kripnografi), mw (Mysql-Waterfall), mv (Mysql-Visual Studio), dn (Data Mining-Naïve Bayes), ws (Waterfall-Sistem Pendukung Keputusan), wsa (Waterfall-Simple Additive Weighting), ss (System Pendukung Keputusan-Simple Additive Weighting), cmp (Customer Relationship Management-Mysql-Php), cp (Codeigniter-Php), cm (Codeigniter-Mysql), es (Expert System Development Life Cycle-Sistem Pakar), am (Augmented Reality-Media Pembelajaran), dp (Data Mining-Padi), dc (Data Mining-C45), mw (Manajemen Bandwidth-Wireless), wp (Web-Php), wm (Web-Mysql), cw (Codeigniter-Waterfall), dan kn (Klasifikasi-Naïve Bayes).

A. Text Preprocessing

Tahap Text Preprocessing ini memiliki beberapa tahapan yang dilakukan seperti case folding, tokenizing, filtering, dan stemming. Hasil dari tahapan ini yaitu sebagai berikut.

Gambar 3 Hasil Preprocessing

Gambar 3 merupakan gambar hasil dari tahapan Case Folding mengubah huruf menjadi huruf kecil atau tanpa huruf kapital dan menghilangkan karakter seperti tanda baca, tahapan Tokenizing pemotongan string berdasarkan tiap kata yang menyusunnya, tahapan Filtering mengambil kata-kata yang penting dari hasil proses Tokenizing, dan tahapan Stemming yaitu proses pengembalian kata dari berbagai bentukan menjadi kata dasar.

B. Rabin Karp

Rabin Karp memiliki beberapa tahapan yang dilakukan seperti K-gram, Hashing, Fingerprint, dan Dice Similarity Coefficient. Hasil dari algoritma ini yaitu sebagai berikut.

Tabel 1 Customer Relationship Management-Mysql-Php

Kode	2 gram	3 gram	4 gram	5 gram	6 gram
cmp1-cmp2	81.00%	49.35%	29.59%	21.53%	17.05%
cmp1-cmp3	81.99%	49.94%	30.21%	21.87%	17.20%
cmp1-cmp4	81.99%	49.94%	30.21%	21.87%	17.20%
cmp1-cmp5	91.02%	82.58%	78.51%	77.09%	75.93%
cmp1-cmp6	93.47%	86.23%	83.82%	81.49%	81.05%
cmp2-cmp3	99.10%	98.23%	97.72%	97.10%	96.84%
cmp2-cmp4	99.10%	98.23%	97.72%	97.10%	96.84%
cmp2-cmp5	90.43%	68.03%	52.63%	47.09%	44.21%
cmp2-cmp6	76.16%	46.40%	27.69%	19.56%	15.42%
cmp3-cmp4	100%	100%	100%	100%	100%
cmp3-cmp5	91.33%	68.36%	52.87%	46.88%	43.68%
cmp3-cmp6	77.23%	47.06%	28.41%	19.96%	15.60%
cmp4-cmp5	91.33%	68.36%	52.87%	46.88%	43.68%
cmp4-cmp6	77.23%	47.06%	28.41%	19.96%	15.60%
cmp5-cmp6	84.44%	68.48%	61.74%	57.91%	56.39%
cmp11-cmp12	84.81%	57.80%	37.56%	27.68%	21.39%
cmp11-cmp13	85.14%	58.45%	38.06%	27.98%	21.58%
cmp11-cmp14	85.14%	58.45%	38.06%	27.98%	21.58%
cmp11-cmp15	92.22%	86.85%	81.97%	79.32%	77.13%
cmp11-cmp16	96.34%	92.14%	88.29%	86.31%	85.47%
cmp12-cmp13	99.16%	98.41%	98.43%	98.19%	98.03%
cmp12-cmp14	99.16%	98.41%	98.43%	98.19%	98.03%
cmp12-cmp15	92.68%	71.12%	55.61%	47.89%	43.13%
cmp12-cmp16	81.31%	56.28%	35.97%	27.04%	21.24%
cmp13-cmp14	100%	100%	100%	100%	100%
cmp13-cmp15	92.97%	71.76%	56.10%	47.95%	42.90%

Kode	2 gram	3 gram	4 gram	5 gram	6 gram
cmp13-cmp16	81.66%	56.99%	36.54%	27.26%	21.47%
cmp14-cmp15	92.97%	71.76%	56.10%	47.95%	42.90%
cmp14-cmp16	81.66%	56.99%	36.54%	27.26%	21.47%
cmp15-cmp16	88.51%	78.53%	69.27%	64.25%	60.80%

Tabel 2 Data Mining-C45

Kode	2 gram	3 gram	4 gram	5 gram	6 gram
dc1-dc2	71.72%	40.99%	25.39%	18.35%	14.72%
dc1-dc3	74.30%	41.83%	22.75%	15.15%	11.31%
dc1-dc4	78.43%	53.77%	39.57%	32.90%	28.81%
dc1-dc5	75.14%	43.66%	23.31%	16.55%	12.45%
dc1-dc6	81.44%	65.27%	53.04%	47.12%	42.16%
dc2-dc3	81.74%	57.18%	44.15%	37.54%	33.02%
dc2-dc4	73.77%	42.12%	28.76%	21.45%	16.87%
dc2-dc5	78.79%	55.56%	40.84%	34.11%	29.58%
dc2-dc6	70.55%	40.21%	24.02%	17.42%	13.32%
dc3-dc4	77.17%	44.65%	27.99%	19.65%	14.28%
dc3-dc5	77.25%	54.57%	35.78%	26.57%	21.88%
dc3-dc6	70.95%	40.39%	23.21%	15.50%	11.33%
dc4-dc5	79.58%	47.34%	28.60%	20.57%	15.71%
dc4-dc6	76.75%	52.22%	37.45%	30.30%	26.33%
dc5-dc6	74.01%	40.99%	20.71%	12.76%	8.25%
dc11-dc12	74.29%	48.36%	30.79%	22.63%	18.09%
dc11-dc13	78.00%	51.75%	32.47%	23.62%	18.11%
dc11-dc14	81.31%	59.85%	45.35%	36.01%	30.70%
dc11-dc15	78.88%	50.58%	29.58%	21.09%	15.44%
dc11-dc16	84.32%	70.00%	58.26%	51.41%	45.98%
dc12-dc13	81.84%	61.52%	45.41%	38.15%	33.13%
dc12-dc14	76.77%	51.73%	36.55%	27.19%	20.77%
dc12-dc15	80.30%	59.96%	45.16%	37.19%	32.18%
dc12-dc16	75.20%	50.61%	32.83%	23.41%	18.06%
dc13-dc14	79.25%	54.50%	37.29%	26.07%	19.46%
dc13-dc15	81.24%	62.11%	43.38%	32.80%	27.07%
dc13-dc16	75.38%	51.36%	34.30%	24.19%	17.46%
dc14-dc15	81.06%	52.90%	34.76%	24.54%	18.45%
dc14-dc16	78.68%	61.69%	46.90%	37.37%	31.71%
dc15-dc16	76.21%	50.19%	29.93%	19.37%	13.43%

Pada kode dc1-dc2 dan dc11-dc12 adalah data abstrak yang sama tetapi memiliki perbedaan tahapan preprocessing, dc1-dc2 melewati semua tahapan preprocessing yaitu tahapan case folding, tokenizing, filtering, dan stemming, sedangkan kode dengan dc11-dc12 hanya melewati tahapan case folding saja, begitu juga dengan semua kode contohnya cmp1-cmp2 dan cmp11-cmp12, cmp1-cmp3 dan cmp11-cmp13 dan seterusnya.

C. Pembahasan

Rata-rata dari setiap kata kunci dan total rata-ratanya sebagai berikut.

Tabel 20. Total rata-rata

Kode		2 gram	3 gram	4 gram	5 gram	6 gram
Android-Kripnografi	Preprocessing	93.57%	88.43%	84.36%	81.66%	79.10%
	Case Folding	93.93%	89.46%	84.37%	80.93%	78.50%
	Total	93.75%	88.95%	84.37%	81.30%	78.80%
Mysql-Waterfall	Preprocessing	81.90%	54.36%	34.91%	25.19%	20.05%
	Case Folding	84.93%	63.25%	43.62%	32.60%	25.42%
	Total	83.42%	58.80%	39.26%	28.90%	22.73%
Mysql-Visual Studio	Preprocessing	83.47%	58.50%	40.18%	32.92%	28.38%
	Case Folding	87.22%	66.82%	50.67%	41.47%	36.36%
	Total	85.35%	62.66%	45.43%	37.19%	32.37%
Data Mining-Naïve Bayes	Preprocessing	78.19%	51.05%	36.99%	29.02%	23.64%
	Case Folding	81.11%	59.14%	43.83%	34.64%	28.78%
	Total	79.65%	55.09%	40.41%	31.83%	26.21%
Waterfall-Sistem Pendukung Keputusan	Preprocessing	84.52%	58.54%	38.57%	29.42%	24.61%
	Case Folding	85.50%	64.66%	46.82%	36.29%	29.90%
	Total	85.01%	61.60%	42.69%	32.85%	27.26%
Waterfall-Simple Additive Weighting	Preprocessing	93.84%	85.68%	78.88%	74.71%	71.43%
	Case Folding	95.05%	90.55%	85.94%	83.48%	81.52%
	Total	94.45%	88.12%	82.41%	79.10%	76.48%
System Pendukung Keputusan-Simple Additive Weighting	Preprocessing	85.16%	58.40%	38.58%	30.02%	25.42%
	Case Folding	86.14%	65.14%	47.36%	37.33%	31.24%
	Total	85.65%	61.77%	42.97%	33.67%	28.33%
Customer Relationship Management-Mysql-Php	Preprocessing	87.72%	68.55%	56.83%	51.75%	49.11%
	Case Folding	90.25%	74.26%	61.80%	55.68%	51.81%
	Total	88.99%	71.41%	59.31%	53.72%	50.46%
Codeigniter-Php	Preprocessing	91.62%	75.02%	65.01%	59.35%	56.81%
	Case Folding	91.87%	80.78%	70.24%	64.04%	60.04%
	Total	91.75%	77.90%	67.63%	61.70%	58.43%
Codeigniter-Mysql	Preprocessing	85.84%	64.00%	48.32%	40.00%	36.05%
	Case Folding	88.65%	71.52%	55.57%	46.69%	41.02%
	Total	87.24%	67.76%	51.95%	43.35%	38.54%
Expert System Development Life Cycle-Sistem Pakar	Preprocessing	85.46%	60.51%	42.49%	34.30%	30.13%
	Case Folding	87.53%	66.47%	49.56%	39.02%	34.48%
	Total	86.49%	63.49%	46.03%	36.66%	32.31%
Augmented Reality-Media Pembelajaran	Preprocessing	86.78%	82.30%	76.79%	72.87%	69.90%
	Case Folding	89.36%	86.17%	82.04%	78.35%	74.89%
	Total	88.07%	84.24%	79.42%	75.61%	72.40%
Data Mining-Padi	Preprocessing	85.40%	73.68%	61.85%	54.72%	50.45%
	Case Folding	88.56%	79.24%	67.67%	60.49%	55.66%
	Total	86.98%	76.46%	64.76%	57.61%	53.06%

Kode		2 gram	3 gram	4 gram	5 gram	6 gram
Data Mining-C45	Preprocessing	76.11%	48.05%	31.70%	24.40%	20.00%
	Case Folding	78.85%	55.81%	38.86%	29.67%	24.00%
	Total	77.48%	51.93%	35.28%	27.03%	22.00%
Manajemen Bandwidth-Wireless	Preprocessing	88.15%	78.49%	72.91%	69.78%	67.71%
	Case Folding	89.25%	79.85%	74.52%	70.23%	68.15%
	Total	88.70%	79.17%	73.72%	70.01%	67.93%
Web-Php	Preprocessing	89.79%	66.77%	49.45%	40.84%	36.26%
	Case Folding	90.36%	73.54%	56.87%	47.15%	40.80%
	Total	90.08%	70.16%	53.16%	43.99%	38.53%
Web-Mysql	Preprocessing	86.64%	62.59%	44.16%	34.89%	30.22%
	Case Folding	88.65%	69.87%	51.90%	41.77%	35.26%
	Total	87.65%	66.23%	48.03%	38.33%	32.74%
Codeigniter-Waterfall	Preprocessing	85.84%	64.00%	48.32%	40.00%	36.05%
	Case Folding	88.65%	71.47%	55.57%	46.69%	40.94%
	Total	87.24%	67.74%	51.95%	43.35%	38.49%
Klasifikasi-Naïve Bayes	Preprocessing	81.74%	65.01%	54.46%	48.19%	44.34%
	Case Folding	82.27%	71.48%	58.46%	50.52%	45.02%
	Total	82.01%	68.25%	56.46%	49.36%	44.68%
Total	Preprocessing	85.88%	66.52%	52.88%	46.00%	42.09%
	Case Folding	87.80%	72.60%	59.25%	51.42%	46.51%
	Total	86.84%	69.56%	56.06%	48.71%	44.30%

Dari data yang digunakan menghasilkan nilai rata-rata total kemiripan 86.84% pada 2 gram, 69.56% pada 3 gram, 56.06% pada 4 gram, 48.71% pada 5 gram, dan 44.30% pada 6 gram. Dimana nilai persentase kemiripan data mulai dari yang terkecil yaitu 70.55% pada 2 gram, 40.21% pada 3 gram, 20.71% pada 4 gram, 12.76% pada 5 gram, dan 8.25% pada 6 gram hingga yang tertinggi 100% pada semua k gram. Dari data yang digunakan dapat dilihat juga bahwa hasil dari Preprocessing (case folding, tokenizing, filtering, dan stemming) dengan hasil dari Preprocessing (case folding), memiliki perbedaan yaitu, hasil dari Preprocessing (case folding, tokenizing, filtering, dan stemming) lebih kecil persentase kemiripannya daripada hasil Preprocessing (case folding), hal ini disebabkan adanya penghilangan kata yang sering digunakan pada tahapan filtering dan adanya perubahan kata pada tahapan stemming.

IV. KESIMPULAN

Pengujian dengan beberapa macam nilai k-gram (2 gram, 3 gram, 4 gram, 5 gram dan 6 gram) dan perbedaan pada preprocessing yaitu dengan menggunakan semua tahapan (case folding, tokenizing, filtering, dan stemming) dan hanya menggunakan tahapan case folding menghasilkan nilai kemiripan yang bermacam – macam, nilai rata-rata total kemiripannya yaitu 86.84% pada 2 gram, 69.56% pada 3 gram, 56.06% pada 4 gram, 48.71% pada 5 gram, dan 44.30% pada 6 gram. Hal ini dapat disimpulkan bahwa dari data yang digunakan masih terlihat adanya tindakan plagiarisme pada abstrak, hal ini didukung dengan adanya data yang memiliki nilai kemiripan hingga 100%.

PENGAKUAN

Makalah ini adalah sebagian dari penelitian Tugas Akhir milik Indra Gunawan dan disponsori oleh Universitas Singaperbangsa Karawang.

DAFTAR PUSTAKA

- [1] A. Millah and S. Nurazizah, “Perbandingan Penggunaan Algoritma Cosinus dan Wu Palmer untuk Mencari Kemiripan Kata dalam,” vol. 2, no. 1, p. 11, 2017.
- [2] A. H. Purba and Z. Situmorang, “Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks,” vol. 02, p. 9, 2017.
- [3] L. Sibarani, M. Magdalena, and A. Dharma, “Analisa Perbandingan Sistem Pendekripsi Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing Dan Algoritma Rabin Karp,” remik, vol. 4, no. 1, p. 69, Oct. 2019, doi: 10.33395/remik.v4i1.10174.
- [4] T. Suryati, Y. Wibisono, and Y. Wihardi, “Aplikasi Deteksi Plagiarisme Dokumen Skripsi dengan Algoritma Rabin-Karp,” vol. 1, no. 2, p. 5, 2018.
- [5] R. Satria, “IDENTIFIKASI BENTUK TINDAK PLAGIAT PADA PENULISAN SKRIPSI MAHASISWA PROGRAM STUDI PENDIDIKAN FISIKA UNSYIAH,” p. 7.

- [6] A. Filcha and M. Hayaty, "Implementasi Algoritma Rabin-Karp untuk Pendekripsi Plagiarisme pada Dokumen Tugas Mahasiswa," *Juita*, vol. 7, no. 1, p. 25, May 2019, doi: 10.30595/juita.v7i1.4063.
- [7] A. Sukmana and A. Sunyoto, "PERBANDINGAN PENGUNAAN STEMMING PADA DETEKSI KEMIRIPAN DOKUMEN MENGGUNAKAN METODE RABIN KARP DAN JACCARD SIMILARITY," p. 6, 2018.
- [8] D. Steveson, H. Agung, and F. Mulia, "APLIKASI PENDETEKSI PLAGIARISME TUGAS DAN MAKALAH PADA SEKOLAH MENGGUNAKAN ALGORITMA RABIN KARP," *J-ALU*, vol. 1, no. 1, May 2018, doi: 10.30813/j-alu.v1i1.1104.G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (*references*)