

KLASIFIKASI BANK SOAL BERDASARKAN KOMPETENSI DASAR MENGGUNAKAN HASIL *SIMILARITY* TERTINGGI ALGORITMA *WINNOWER*

Azis Suroni¹⁾, Endang Setyati²⁾

¹⁾Program Studi Teknologi Informasi

²⁾Institut Sains dan Teknologi Terpadu Surabaya

Email : unisba.4215@gmail.com ¹⁾, endang@stts.edu ²⁾

Abstrak : Dunia pendidikan di Indonesia perlu menyiapkan siswa untuk menghadapi tantangan abad 21 yang semakin kompleks. Mereka harus mampu berpikir kritis, kreatif, inovatif dan mampu memecahkan masalah. Dalam hal tersebut maka perlu dilakukan pengukuran capaian pembelajaran siswa dengan melakukan penilaian baik secara tertulis maupun lisan. Dalam penilaian secara tertulis, seorang pendidik diharapkan memiliki banyak bank soal yang beragam sehingga bisa mengukur capaian kompetensi siswa secara maksimal. Banyaknya bank soal terkadang membuat pendidik/ guru kesulitan untuk memilih soal untuk mengukur sejauh mana kesuksesan siswa dalam mencapai kompetensi tertentu. Dengan Algoritma *Winnower* dilakukan proses ekstraksi fitur fingerprint pada soal yang ada pada bank soal kemudian dibandingkan dengan fitur fingerprint dari kompetensi dasar yang sudah ada dengan menggunakan Cosine Similarity. Hasil similarity tertinggi digunakan untuk menentukan soal tersebut masuk kategori kompetensi dasar yang mana. Penelitian ini juga digunakan untuk mencari kombinasi terbaik dari algoritma *winnower* dalam melakukan ekstraksi fitur fingerprint yang digunakan sebagai perbandingan dua buah teks. Penelitian ini dilakukan dengan mengubah nilai n , w dan p pada proses *winnower*. Penelitian dilakukan 3 kali ujicoba dengan nilai $n = \{2,3,4\}$, $w = \{2,3,4\}$ dan nilai p konstan = 2. Dari hasil penelitian yang dilakukan diperoleh kombinasi terbaik dengan nilai $n=2$, $w=2$ dan $p=2$ dengan hasil klasifikasi similarity tertinggi sebesar 85%. Dengan adanya penelitian ini diharapkan dapat memberi kemudahan bagi pendidik untuk memilih soal pada bank soal dan memberikan kemudahan untuk mengetahui capaian siswa pada kompetensi dasar tertentu.

Kata Kunci–Penilaian, Bank Soal, kompetensi dasar, *winnower*, Cosine Similarity

Abstract: The world of education in Indonesia needs to prepare students to face the increasingly complex challenges of the 21st century. They must be able to think critically, creatively, innovatively, and able to solve problems. In this case, it is necessary to measure student learning achievement by making an assessment both in writing and orally. In a written assessment, an educator is expected to have many diverse question banks so that they can measure the achievement of student competencies to the fullest. The number of question banks sometimes makes it difficult for educators/teachers to sort out questions to measure the extent to which students are successful in achieving certain competencies. With the *Winnower* Algorithm, the fingerprint feature extraction process is carried out on the questions in the question bank and then compared with the fingerprint features of the existing basic competencies using Cosine Similarity. The results of the highest similarity are used to determine which questions fall into the category of basic competencies. This research is also used to find the best combination of *winnower* algorithms in performing fingerprint feature extraction which is used as a comparison of two texts. This research was conducted by changing the values of n , w , and p in the *winnower* process. The research was carried out 3 times with a value of $n = \{2,3,4\}$, $w = \{2,3,4\}$ and a constant p value = 2. From the results of the research conducted, the best combination was obtained with a value of $n=2$, $w=2$, and $p=2$ with the highest similarity classification result of 85%. With this research, it is hoped that it will make it easier for educators to sort out questions in the question bank and make it easier to find out student achievements in certain basic competencies.

Keywords–Assessment, Question Bank, basic competence, *winnower*, Cosine Similarity

I. PENDAHULUAN

DUNIA pendidikan di Indonesia perlu menyiapkan peserta didik untuk menghadapi tantangan abad 21 yang semakin kompleks. Peserta didik harus mampu mengembangkan kecakapan esensial di abad ini. Peserta didik juga dituntut mampu berpikir kritis, kreatif dan inovatif dalam menyelesaikan sebuah permasalahan. Untuk mengetahui kemampuan siswa dalam menghadapi tantangan abad ini, perlu dilakukan pengukuran capaian pembelajaran siswa oleh pendidik atau pun satuan pendidikan. Penilaian yang dilakukan pendidik atau pun satuan pendidikan diharapkan bukan hanya mengukur aspek pengetahuan saja, namun juga aspek sikap dan keterampilan. Bentuk penilaian tidak

hanya menggunakan terterulis namun juga menggunakan bentuk penilaian lain seperti tes lisan dan penugasan [1].

Dalam hal penilaian capaian pembelajaran siswa, pendidik memiliki dokumen bank soal yang akan/telah diujikan ke peserta didik. Soal-soal tersebut disusun melalui analisis kompetensi dasar yang hendak dicapai oleh peserta didik dalam pembelajaran.

Melakukan penyesuaian antara kompetensi dasar dengan soal yang dibuat bisa dilakukan dengan memanfaatkan Algoritma *Winnowing*. Algoritma *Winnowing* melakukan perbandingan kemiripan antara dua dokumen yang ada [2].

Berdasarkan dengan paparan tersebut, penulis melakukan penelitian tentang klasifikasi soal berdasarkan kompetensi dasar yang sesuai dengan kurikulum yang berlaku dengan menggunakan Algoritma *Winnowing*, sehingga memberi kemudahan bagi pendidik dalam melakukan penyusunan soal dalam proses penilaian terhadap capaian pembelajaran siswa.

II. TINJAUAN PUSTAKA

A. Klasifikasi

Klasifikasi adalah proses untuk mengelompokkan sejumlah data ke dalam kelas-kelas tertentu yang sudah diberikan berdasarkan kesamaan sifat dan pola yang terdapat dalam data-data tersebut. Secara umum, proses klasifikasi dimulai dengan diberikannya sejumlah data yang menjadi acuan untuk membuat aturan klasifikasi data. Data-data ini biasa disebut dengan *training sets*, dari *training sets* tersebut kemudian dibuat suatu model untuk mengklasifikasikan data. Model tersebut kemudian digunakan sebagai acuan untuk mengklasifikasikan data-data yang belum diketahui kelasnya yang disebut dengan *test set* atau data latih [3].

Klasifikasi (*classification*) adalah metode *data mining* yang dapat digunakan untuk proses pencarian sekumpulan model (fungsi) yang dapat menjelaskan dan membedakan kelas-kelas data atau konsep, yang tujuannya supaya model tersebut dapat digunakan memprediksi objek kelas yang labelnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang akan muncul di masa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data [4].

B. Kompetensi Dasar

Mengacu pada Permendikbud No 24 Tahun 2016, Kompetensi Dasar (KD) adalah kemampuan minimal yang harus dikuasai peserta didik setelah mempelajari materi pelajaran tertentu sesuai dengan kurikulum yang digunakan oleh satuan pendidikan.

Kompetensi dasar berisi sejumlah kemampuan yang harus dikuasai peserta didik dalam mata pelajaran tertentu, sebagai rujukan penyusunan indikator kompetensi dalam suatu pelajaran. Di dalam setiap rumusan Kompetensi Dasar, terdapat unsur kemampuan berpikir yang dinyatakan dalam kata kerja dan materi [5].

Kompetensi Dasar berisi sikap, pengetahuan, dan ketrampilan yang bersumber pada kompetensi inti yang harus dikuasai peserta didik. Kompetensi tersebut dikembangkan dengan memperhatikan karakteristik peserta didik, kemampuan awal, dan ciri suatu mata pelajaran.

C. Algoritma *Winnowing*

Algoritma *Winnowing* merupakan algoritma yang digunakan dalam deteksi kesamaan menggunakan fungsi *hashing*. *Winnowing* adalah algoritma yang digunakan untuk melakukan proses *document fingerprinting*. Algoritma *Winnowing* melakukan penghitungan nilai-nilai *hash* dari setiap *k-gram*, untuk mencari nilai *hash* selanjutnya digunakan fungsi *rolling hash*. Kemudian dibentuk *window* dari nilai-nilai *hash* tersebut. Dalam setiap *window* dipilih nilai *hash* minimum. Jika ada lebih dari satu *hash* dengan nilai minimum, dipilih nilai *hash* yang paling kanan. Kemudian semua nilai *hash* terpilih disimpan untuk dijadikan *fingerprint* dari suatu dokumen. *Fingerprint* ini yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukkan [6].

N-gram adalah metode berbasis karakter. Langkah pertama dari *Winnowing* adalah pembentukan *n-gram*. Metode ini memiliki masukan *n*, dimana nilai ini akan diubah menjadi jumlah potongan atau substring dari teks atau kata-kata yang terbentuk. Pada proses ini membutuhkan *substring* karakter semua *n* struktur kalimat. Pengambilan dilakukan dari awal hingga akhir dokumen atau teks. Jika nilai *n* terlalu besar, kata yang didapat akan panjang, hal ini menyebabkan kata-kata penting tidak tertangkap. Namun, jika *n* terlalu kecil akan ada banyak potongan kata. Hal ini akan menjadi sulit untuk membedakan antar kata. Dengan alasan tersebut, nilai yang sesuai dari *n* harus dipilih yang merupakan nilai minimum yang dapat membedakan *coincidental similarity*. *N-gram* dibutuhkan dalam Algoritma *Winnowing* sebagai masukan untuk mendapatkan kata-kata tertentu dari teks [7]. Contoh proses *N-Gram* terdapat pada Tabel 1 sebagai berikut:

Tabel 1. Proses *N-Gram*

Hasil Preprocessing	N-Gram (k=2)
analismemoridasarkarakteristiksisistemmem orilokasikapasitascepatcaraaksestipefisik	an, na, al, li, is, sm, me, em, mo, or, ri, id, da, as, sa, ar, rk, ka, ar, ra, ak, kt, te, er, ri, is, st, ti, ik, ks, si, is, st, te, em, mm, me, em, mo, or, ri, il, lo, ok, ka, as, si, ik, ka, ap, pa, as, si, it, ta, as, sc, ce, ep, pa, at, tc, ca, ar, ra, aa, ak, ks, se, es, st, ti, ip, pe, ef, fi, is, si, ik

Hashing adalah teknik mengubah karakter string yang mewakili nilai aslinya. Nilai dari nilai hash atau hash yang dari algoritma *winnowing* mewakili *fingerpint* dari sebuah teks. Nilai yang diperoleh dari perhitungan nilai ASCII dari setiap karakter. Dalam hal ini, karakter dalam pertanyaan adalah karakter yang dibentuk dari *N-gram*.

Tipe hashing yang digunakan dalam proses ini menggunakan algoritma Rabin-Karp atau biasa dikenal *rolling hash* [8]. Algoritma ini digunakan untuk menyelesaikan masalah yang berhubungan dengan pencocokan string. *Rolling hash* digunakan untuk mereduksi perulangan ketika menghitung nilai *hash*, membuat proses lebih efektif [9]. Fungsi matematis dari *rolling hash* adalah sebagai berikut :

$$H_{C_2...C_{K+1}} = (H_{C_2...C_K} - C_1 * b^{(k-1)}) * b + H_{(C_{K+1})} \quad (1)$$

Keterangan :

- c : nilai ASCII
- b : basis (bilangan prima)
- k : jumlah karakter

Contoh menghitung nilai *hash* dari kata “bunga” dengan nilai bilangan prima *b* = 2 adalah sebagai berikut:

$$\begin{aligned} H_{(bunga)} &= b * 2^4 + u * 2^3 + n * 2^2 + g * 2^1 + a * 2^0 \\ &= 98 * 16 + 117 * 8 + 110 * 4 + 103 * 2 + 97 * 1 \\ &= 1568 + 936 + 440 + 206 + 97 \\ &= 3247 \end{aligned}$$

Contoh hasil proses hashing dapat dilihat pada Tabel 2. berikut:

Tabel 2. Hasil Proses *Hashing*

N-Gram (k=2)	Hashing (p =2)
an, na, al, li, is, sm, me, em, mo, or, ri, id, da, as, sa, ar, rk, ka, ar, ra, ak, kt, te, er, ri, is, st, ti, ik, ks, si, is, st, te, em, mm, me, em, mo, or, ri, il, lo, ok, ka, as, si, ik, ka, ap, pa, as, si, it, ta, as, sc, ce, ep, pa, at, tc, ca, ar, ra, aa, ak, ks, se, es, st, ti, ip, pe, ef, fi, is, si, ik	[608,634,604,642,650,678,638,622,658,672,666,620,594,6 18,654,616,670,622,616,650,602,660,666,632,666,650, 692,674,634,658,670,650,692,666,622,654,638,622,658,6 72,666,636,654,658,622,618,670,634,622,612,642,618,67 0,652,658,618,658,598,628,642,620,662,590,616,650,582, 602,658,662,634,692,674,644,650,608,618,650,670,634]

Windowing adalah proses dari algoritma *winnowing* untuk menyimpan hasil dari *rolling hash* yang telah dihasilkan sebelumnya. Nilai *hash* dikelompokkan ke dalam *window w*, dimana nilai ini perlu

ditentukan sebelumnya. Hasil dari proses *windowing* ini adalah serangkaian nilai hasil hasing dari semua w yang telah dikelompokkan [10]. Lihat Tabel 3.

Tabel 3. Hasil *Windowwing*

N-Gram (k=2)	Hashing (p =2)	Windowwing (w=2)
an, na, al, li, is, sm, me, em, mo, or, ri, id, da, as, sa, ar, rk, ka, ar, ra, ak, kt, te, er, ri, is, st, ti, ik, ks, si, is, st, te, em, mm, me, em, mo, or, ri, il, lo, ok, ka, as, si, ik, ka, ap, pa, as, si, it, ta, as, sc, ce, ep, pa, at, tc, ca, ar, ra, aa, ak, ks, se, es, st, ti, ip, pe, ef, fi, is, si, ik	[608,634,604,642,650,67 8,638,622,658,672,666,6 20,594,618,654,616,670, 622,616,650,602,660,66 6,632,666,650,692,674,6 34,658,670,650,692,666, 622,654,638,622,658,67 2,666,636,654,658,622,6 18,670,634,622,612,642, 618,670,652,658,618,65 8,598,628,642,620,662,5 90,616,650,582,602,658, 662,634,692,674,644,65 0,608,618,650,670,634]	[[608,634],[634,604],[604,642],[642,650],[650,678],[678,6 38],[638,622],[622,658],[658,672],[672,666],[666,620],[62 0,594],[594,618],[618,654],[654,616],[616,670],[670,622],[622,616],[616,650],[650,602],[602,660],[660,666],[666,63 2],[632,666],[666,650],[650,692],[692,674],[674,634],[634, 658],[658,670],[670,650],[650,692],[692,666],[666,622],[6 22,654],[654,638],[638,622],[622,658],[658,672],[672,666]],[666,636],[636,654],[654,658],[658,622],[622,618],[618,6 70],[670,634],[634,622],[622,612],[612,642],[642,618],[61 8,670],[670,652],[652,658],[658,618],[618,658],[658,598],[598,628],[628,642],[642,620],[620,662],[662,590],[590,61 6],[616,650],[650,582],[582,602],[602,658],[658,662],[662, 634],[634,692],[692,674],[674,644],[644,650],[650,608],[6 08,618],[618,650],[650,670],[670,634]]

Proses utama dari Algoritma *Winnowing* adalah pemilihan dari *fingerpint*. Proses ini bisa dikerjakan dengan mencari nilai *hash* minimum dari setiap window. Jika dalam window ada nilai minimum lebih dari satu, *hash* yang berada di paling kanan akan dipilih sebagai *fingerpint*. Nilai-nilai ini akan dipasang ke dalam *fingerpint*. Nilai dari *fingerpint* ini akan digunakan sebagai referensi dalam menghitung kemiripan antara dua teks tersebut. Lihat Tabel 4.

Tabel 4. Proses *Fingerprinting*

Windowwing (w=2)	Fingerprinting
[[608,634],[634,604],[604,642],[642,650],[650,678],[678,638],[638,622],[6 22,658],[658,672],[672,666],[666,620],[620,594],[594,618],[618,654],[654, 616],[616,670],[670,622],[622,616],[616,650],[650,602],[602,660],[660,66 6],[666,632],[632,666],[666,650],[650,692],[692,674],[674,634],[634,658],[658,670],[670,650],[650,692],[692,666],[666,622],[622,654],[654,638],[63 8,622],[622,658],[658,672],[672,666],[666,636],[636,654],[654,658],[658,6 22],[622,618],[618,670],[670,634],[634,622],[622,612],[612,642],[642,618]],[618,670],[670,652],[652,658],[658,618],[618,658],[658,598],[598,628],[6 28,642],[642,620],[620,662],[662,590],[590,616],[616,650],[650,582],[582, 602],[602,658],[658,662],[662,634],[634,692],[692,674],[674,644],[644,65 0],[650,608],[608,618],[618,650],[650,670],[670,634]]	[608,604,604,642,650,638,622, 622,658,666,620,594,594,618,6 16,616,622,616,616,602,602,66 0,632,632,650,650,674,634,634 ,658,650,650,666,622,622,638, 622,622,658,666,636,636,654,6 22,618,618,634,622,612,612,61 8,618,652,652,618,618,598,598 ,628,620,620,590,590,616,582, 582,602,658,634,634,674,644,6 44,608,608,618,650,634]

D. Cosine Similarity

Cosine similarity digunakan untuk mengukur dua *finite-dimensional vector* dari dimensi yang sama yang mana masing-masing vector merepresentasikan sebuah dokumen. Untuk contoh, jika ada sebuah string s1 = "the sun rises in the east" dan string s2 = "the sun in the sky is bright", maka kumpulan term dari s1 dan s2 adalah T = {the, sun, rises, in, east, sky, is, bright}. Kemudian kita menghitung frekuensi term dari string s1 dan s2. Vector yang dihasilkan adalah $\vec{v}_{s1} = (2, 1, 1, 1, 1, 0, 0, 0)$ dan $\vec{v}_{s2} = (2, 1, 0, 1, 0, 1, 1, 1)$.

Ketika vector dari dokumen sudah terbentuk, kita hitung kemiripannya dengan menggunakan formulasi cosine similarity berikut ;

$$\text{Cos } \theta = \frac{\vec{v}_{a1} \cdot \vec{v}_{a2}}{|\vec{v}_{a1}| \times |\vec{v}_{a2}|} \quad (2)$$

Jika nilai Cos θ mendekati 1, maka kemiripan dokumen semakin besar.

III. METODE PENELITIAN

Pada penelitian ini ada beberapa tahap yang dilakukan antara lain :

A. Pengumpulan Dataset

Dataset dari penelitian ini adalah soal Sistem Komputer pada kompetensi keahlian SMK Teknik Komputer dan Jaringan yang terdiri dari 50 soal pilihan ganda yang dibagi menjadi data training 30 soal dan data testing sebanyak 20 soal, yang dibandingkan dengan kompetensi dasar dari mata pelajaran Sistem Komputer pada kompetensi keahlian SMK Teknik Komputer dan jaringan sesuai dengan kurikulum yang berlaku saat ini yang terdiri dari 10 kompetensi dasar.

B. Preprocessing

Pada tahap ini, data set dilakukan proses *casefolding* yaitu perubahan huruf kecil semua, *stopword removal* yaitu menghilangkan kata-kata yang tidak penting, dan *stemming*. *Stemming* adalah pengembalian kata ke dalam bentuk kata dasar. Contoh hasil *preprocessing* terdapat pada Tabel 5 berikut:

Tabel 5. Hasil Preprocessing Kompetensi Dasar

Kode kd	KD	Preprocessing
SK3.1	Memahami sistem bilangan (Desimal, Biner, Heksadesimal)	pahamsistembilangdesimalbinerheksadesimal
SK3.2	Menganalisis relasi logika dasar, kombinasi dan sekuensial (NOT, AND, OR); (NOR, NAND, EXOR, EXNOR); (Flip Flop, counter)	analisisrelasilogikadasarkombinasidansekuensi alnotandornandorexnorflipflopcounter

C. Winnowing Process

Setelah dilakukan *process preprocessing*, data selanjutnya akan diproses menggunakan algoritma winnowing sehingga didapatkan nilai dari *fingerpint* data tersebut. Proses *winnowing* terdiri dari pembentukan nilai *N-gram*, *hashing*, *windowing* dan *fingerpinting*.

Contoh hasil proses pembentukan *N-gram* terdapat pada Tabel 6 berikut ini:

Tabel 6. Pembentukan N-gram dengan nilai k = 2

Kode kd	KD	N-gram
SK3.1	Memahami sistem bilangan (Desimal, Biner, Heksadesimal)	pa, ah, ha, am, ms, si, is, st, te, em, mb, bi, il, la, an, ng, gd, de, es, si, im, ma, al, lb, bi, in, ne, er, rh, he, ek, ks, sa, ad, de, es, si, im, ma, al
SK3.2	Menganalisis relasi logika dasar, kombinasi dan sekuensial (NOT, AND, OR); (NOR, NAND, EXOR, EXNOR); (Flip Flop, counter)	an, na, al, li, is, sr, re, el, la, as, si, il, lo, og, gi, ik, ka, ad, da, as, sa, ar, rk, ko, om, mb, bi, in, na, as, si, id, da, an, ns, se, ek, ku, ue, en, ns, si, ia, al, ln, no, ot, ta, an, nd, do, or, rn, no, or, rn, na, an, nd, de, ex, xo, or, re, ex, xn, no, or, rf, fl, li, ip, pf, fl, lo, op, pc, co, ou, un, nt, te, er

Hasil dari proses *N-gram* dilanjutkan pada proses *hashing*, yakni mencari nilai *hash* dengan menggunakan algoritma *rolling hash* pada persamaan 1. Input *hash* didapatkan dari data *n-gram* yang sudah terbentuk. Contoh hasil proses hashing dengan nilai $p = 5$ dapat dilihat pada Tabel 7 berikut ini:

Tabel 7. Hashing dengan nilai p = 5

Kode kd	N-gram	Hash Value
SK3.1	pa, ah, ha, am, ms, si, is, st, te, em, mb, bi, il, la, an, ng, gd, de, es, si, im, ma, al, lb, bi, in, ne, er, rh, he, ek, ks, sa, ad, de, es, si, im, ma, al	[3285,2945,3085,2970,3300,3400,3200,3455,3405,3070,3215,2975,3165,3185,2975,3265,3075,3005,3100,3400,3170,3210,2965,3190,2975,3175,3255,3095,3370,3105,3060,3250,3360,2925,3005,3100,3400,3170,3210,2965]
SK3.2	an, na, al, li, is, sr, re, el, la, as, si, il, lo, og, gi, ik, ka, ad, da, as, sa, ar, rk, ko, om, mb, bi, in, na, as, si, id, da, an, ns, se, ek, ku, ue, en, ns, si, ia, al, ln, no, ot, ta, an, nd, do, or,	[2975,3235,2965,3225,3200,3445,3355,3065,3185,3000,3400,3165,3255,3290,3100,3160,3160,2925,2985,3000,3360,2995,3385,3230,3320,3215,2975,3175,3235,3000,3400,3125,2985,2975,3325,3380,3060,3260,3430,3075,3325,3400,3110,2965,3250,3305,3355,3385,2975,3250,3055,3345]

Kode kd	N-gram	Hash Value
	rn, no, or, rn, na, an, nd, de, ex, xo,	,3400,3305,3345,3400,3235,2975,3250,3005,3125,3555,3
	or, re, ex, xn, no, or, rf, fl, li, ip, pf,	345,3355,3125,3550,3305,3345,3360,3090,3225,3185,331
	fl, lo, op, pc, co, ou, un, nt, te, er	0,3090,3255,3335,3295,3030,3360,3475,3330,3405,3095]

Setelah diperoleh nilai *hash*, selanjutnya data akan dilakukan proses *windowing*. Proses *windowing* ini yang membedakan antara algoritma *winnowing* dengan algoritma *fingerprint* dan penentuan *fingerprinting*. Contoh hasil *windowing* dengan nilai $w = 2$ dapat dilihat pada Tabel 8 berikut ini:

Tabel 8. Proses Windowing Dengan Nilai $w = 2$

Kode kd	Hash Value	Windowing
SK3.1	[3285,2945,3085,2970,3300,3400,3200,3455,3405,3070,3215,2975,3165,3185,2975,3265,3075,3005,3100,3400,3170,3210,2965,3190,2975,3175,3255,3095,3370,3105,3060,3250,3360,2925,3005,3100,3400,3170,3210,2965]	[[3285,2945],[2945,3085],[3085,2970],[2970,3300],[3300,3400],[3400,3200],[3200,3455],[3455,3405],[3405,3070],[3070,3215],[3215,2975],[2975,3165],[3165,3185],[3185,2975],[2975,3265],[3265,3075],[3075,3005],[3005,3100],[3100,3400],[3400,3170],[3170,3210],[3210,2965],[2965,3190],[3190,2975],[2975,3175],[3175,3255],[3255,3095],[3095,3370],[3370,3105],[3105,3060],[3060,3250],[3250,3360],[3360,2925],[2925,3005],[3005,3100],[3100,3400],[3400,3170],[3170,3210],[3210,2965]]
SK3.2	[2975,3235,2965,3225,3200,3445,3355,3065,3185,3000,3400,3165,3255,3290,3100,3160,3160,2925,2985,3000,3360,2995,3385,3230,3320,3215,2975,3175,3235,3000,3400,3125,2985,2975,3325,3380,3060,3260,3430,3075,3325,3400,3110,2965,3250,3305,3355,3385,2975,3250,3055,3345,3400,3305,3345,3400,3235,2975,3250,3005,3125,3555,3345,3355,3125,3550,3305,3345,3360,3090,3225,3185,3310,3090,3255,3335,3295,3030,3360,3475,3330,3405,3095]	[[2975,3235],[3235,2965],[2965,3225],[3225,3200],[3200,3445],[3445,3355],[3355,3065],[3065,3185],[3185,3000],[3000,3400],[3400,3165],[3165,3255],[3255,3290],[3290,3100],[3100,3160],[3160,3160],[3160,2925],[2925,2985],[2985,3000],[3000,3360],[3360,2995],[2995,3385],[3385,3230],[3230,3320],[3320,3215],[3215,2975],[2975,3175],[3175,3235],[3235,3000],[3000,3400],[3400,3125],[3125,2985],[2985,2975],[2975,3325],[3325,3380],[3380,3060],[3060,3260],[3260,3430],[3430,3075],[3075,3325],[3325,3400],[3400,3110],[3110,2965],[2965,3250],[3250,3305],[3305,3355],[3355,3385],[3385,2975],[2975,3250],[3250,3055],[3055,3345],[3345,3400],[3400,3305],[3305,3345],[3345,3400],[3400,3235],[3235,2975],[2975,3250],[3250,3005],[3005,3125],[3125,3555],[3555,3345],[3345,3355],[3355,3125],[3125,3550],[3550,3305],[3305,3345],[3345,3360],[3360,3090],[3090,3225],[3225,3185],[3185,3310],[3310,3090],[3090,3255],[3255,3335],[3335,3295],[3295,3030],[3030,3360],[3360,3475],[3475,3330],[3330,3405],[3405,3095]]

Dari hasil proses *windowing*, diambil nilai *fingerprint* terkecil pada masing-masing *window*, dan disusun menjadi sebuah dokumen *fingerprint* yang nantinya akan dibandingkan antar dokumen *fingerprint* tersebut sehingga diperoleh suatu kemiripan. Contoh hasil *fingerprinting* dapat dilihat pada Tabel 9 berikut ini:

Tabel 9. Proses Fingerprinting

Kode kd	Windowing	Fingerprint
SK3.1	[[3285,2945],[2945,3085],[3085,2970],[2970,3300],[3300,3400],[3400,3200],[3200,3455],[3455,3405],[3405,3070],[3070,3215],[3215,2975],[2975,3165],[3165,3185],[3185,2975],[2975,3265],[3265,3075],[3075,3005],[3005,3100],[3100,3400],[3400,3170],[3170,3210],[3210,2965],[2965,3190],[3190,2975],[2975,3175],[3175,3255],[3255,3095],[3095,3370],[3370,3105],[3105,3060],[3060,3250],[3250,3360],[3360,2925],[2925,3005],[3005,3100],[3100,3400],[3400,3170],[3170,3210],[3210,2965]]	[2945,2945,2970,2970,3300,3200,3200,3405,3070,3070,2975,2975,3165,2975,2975,3075,3005,3005,3100,3170,3170,2965,2965,2975,2975,3175,3095,3095,3105,3060,3060,3250,2925,2925,3005,3100,3170,3170,2965]
SK3.2	[[2975,3235],[3235,2965],[2965,3225],[3225,3200],[3200,3445],[3445,3355],[3355,3065],[3065,3185],[3185,3000],[3000,3400],[3400,3165],[3165,3255],[3255,3290],[3290,3100],[3100,3160],[3160,2925],[2925,2985],[2985,3000],[3000,3360],[3360,2995],[2995,3385],[3385,3230],[3230,3320],[3320,3215],[3215,2975],[2975,3175],[3175,3235],[3235,3000],[3000,3400],[3400,3125],[3125,2985],[2985,2975],[2975,3325],[3325,3380],[3380,3060],[3060,3260],[3260,3430],[3430,3075],[3075,3325],[3325,3400],[3400,3110],[3110,2965],[2965,3250],[3250,3305],[3305,3355],[3355,3385],[3385,2975],[2975,3250],[3250,3055],[3055,3345],[3345,3400],[3400,3305],[3305,3345],[3345,3400],[3400,3235],[3235,2975],[2975,3250],[3250,3005],[3005,3125],[3125,3555],[3555,3345],[3345,3355],[3355,3125],[3125,3550],[3550,3305],[3305,3345],[3345,3360],[3360,3090],[3090,3225],[3225,3185],[3185,3310],[3310,3090],[3090,3255],[3255,3335],[3335,3295],[3295,3030],[3030,3360],[3360,3475],[3475,3330],[3330,3405],[3405,3095]]	[2975,2965,2965,3200,3200,3355,3065,3065,3000,3000,3165,3165,3255,3100,3100,3160,2925,2925,2985,3000,2995,2995,3230,3230,3215,2975,2975,3175,3000,3000,3125,2985,2975,2975,3325,3060,3060,3260,3075,3075,3325,3110,2965,2965,3250,3305,3355,2975,2975,3055,3055,3345,3305,3305,

Kode kd	Windowing	Fingerprint
	2975],[2975,3175],[3175,3235],[3235,3000],[3000,3400],[3400,3125],[3125,2985],[2985,2975],[2975,3325],[3325,3380],[3380,3060],[3060,3260],[3260,3430],[3430,3075],[3075,3325],[3325,3400],[3400,3110],[3110,2965],[2965,3250],[3250,3305],[3305,3355],[3355,3385],[3385,2975],[2975,3250],[3250,3055],[3055,3345],[3345,3400],[3400,3305],[3305,3345],[3345,3400],[3400,3235],[3235,2975],[2975,3250],[3250,3005],[3005,3125],[3125,3555],[3555,3345],[3345,3355],[3355,3125],[3125,3550],[3550,3305],[3305,3345],[3345,3360],[3360,3090],[3090,3225],[3225,3185],[3185,3310],[3310,3090],[3090,3255],[3255,3335],[3335,3295],[3295,3030],[3030,3360],[3360,3475],[3475,3330],[3330,3405],[3405,3095]]	3345,3235,2975,2975,3005,3005,3125,3345,3345,3125,3125,3305,3305,3345,3090,3090,3185,3185,3090,3090,3255,3295,3030,3030,3360,3330,3330,3095]

IV. HASIL PERCOBAAN

Pada penelitian ini, difokuskan pada percobaan perubahan nilai bilangan prima dalam proses winnowing. Dari penelitian ini diharapkan diperoleh kombinasi nilai parameter yang cocok untuk melakukan klasifikasi soal dengan kompetensi dasar dengan menggunakan hasil similarity tertinggi algoritma winnowing. Hasil klasifikasi oleh sistem dibandingkan dengan hasil klasifikasi oleh guru profesional sehingga didapatkan akurasi dari algoritma winnowing dalam melakukan proses klasifikasi soal dengan kompetensi dasar. Tabel 10 merupakan data tes dari soal yang diujikan pada penelitian ini.

Tabel 10. Data tes Soal Sistem Komputer

Kode	Mapel	Soal
SK1	Sistem Komputer	Sifat gerbang logika NOR adalah...
SK2	Sistem Komputer	Transistor merupakan komponen dasar dalam rangkaian elektronika multitermal dengan 3 buah terminal. Ada 3 jenis transistor salah satunya adalah
SK3	Sistem Komputer	Hasil penyandian decimal menjadi biner dalam format BCD akan diproses oleh CPU. Nilai dalam format BCD yang dihasilkan oleh CPU tidak dapat dibaca oleh pengguna sehingga perlu diubah dahulu dalam bentuk decimal dengan menggunakan rangkaian. Rangkaian tersebut adalah
SK4	Sistem Komputer	Sistem bilangan mengenal dua posisi angka, yakni
SK5	Sistem Komputer	Satuan kapasitas memori komputer dari 1 GB (gigabyte) berapa KB (megabyte) ?
SK6	Sistem Komputer	Berikut dibawah ini adalah macam-macam sistem bilangan kecuali ?
SK7	Sistem Komputer	Bilangan biner yaitu bilangan yang berbasis 2 dan bilangan desimal adalah bilangan yang berbasis
SK8	Sistem Komputer	Ada 3 gerbang logika dasar yaitu gerbang AND, OR, NOT. Output akan belogika 0 Jika semua input berlogika 0, dan output akan berlogika 1 jika salah satu input berlogika 1. Logika tersebut merupakan pernyataan logika untuk gerbang logika
SK9	Sistem Komputer	Perbedaan RS flip flop dibandingkan dJK flip flop terletak pada
SK10	Sistem Komputer	Perbedaan Rs flip flop dan D flip flop terdapat tambahan gerbang
SK11	Sistem Komputer	Gerbang logika yang digunakan untuk merangkai D flip flop kecuali
SK12	Sistem Komputer	Suatu rangkaian penjumlahan sistem bilangan biner yang paling sederhana adalah
SK13	Sistem Komputer	Dibawah ini yang bukan merupakan struktur utama di dalam sebuah komputer adalah
SK14	Sistem Komputer	Yang bukan merupakan fungsi utama komputer adalah
SK15	Sistem Komputer	Penemu dari Arsitektur Von Noumann adalah
SK16	Sistem Komputer	Berikut ini yang termasuk kedalam bagian utama dari Arsitektur Von Noumann adalah
SK17	Sistem Komputer	Rangkaian adder yang hanya menjumlahkan dua bit disebut
SK18	Sistem Komputer	Rangkaian adder yang hanya menjumlahkan tiga bit disebut
SK19	Sistem Komputer	Rangkaian adder yang menjumlahkan banyak bit disebut
SK20	Sistem Komputer	Apa nama alat yang digunakan untuk dapat mengembalikan proses encoding sehingga dapat melihat atau menerima informasi aslinya

Soal-soal tersebut akan diklasifikasikan sesuai dengan Kompetensi Dasar mata pelajaran Sistem Komputer yang ada pada Tabel 11 sebagai berikut:

Tabel 11. Kompetensi Dasar Pengetahuan Sistem Komputer

Kode KD	Kompetensi Dasar
3.1.	Memahami sistem bilangan (Desimal, Biner, Heksadesimal)
3.2.	Menganalisis relasi logika dasar, kombinasi dan sekuensial (NOT, AND, OR); (NOR, NAND, EXOR, EXNOR); (Flip Flop, counter)
3.3.	Menerapkan operasi logika Aritmatik (Half-Full Adder, Ripple Carry Adder)
3.4.	Mengklasifikasikan rangkaian Multiplexer, Decoder, Register
3.5.	Menerapkan elektronika dasar (kelistrikan, komponen elektronika dan skema rangkaian elektronika)
3.6.	Menerapkan dasar dasar mikrokontroler
3.7.	Menganalisis blok diagram dari sistem mikro komputer (arsitektur komputer)
3.8.	Mengevaluasi Perangkat Eksternal / Peripheral
3.9.	Menganalisis memori berdasarkan karakteristik sistem memori (lokasi, kapasitas, kecepatan, cara akses, tipe fisik)
3.10.	Menganalisa Struktur CPU dan fungsi CPU

Data tes soal dilakukan ujicoba pada system. Ujicoba pertama pada mata pelajaran Sistem Komputer dilakukan dengan menggunakan nilai parameter $k = 2$, $p = 2$ dan $w = 2$ karena pada beberapa penelitian sebelumnya [12] diperoleh hasil klasifikasi terbaik dengan nilai parameter tersebut. Berdasarkan hasil uji coba system, diperoleh hasil sebagaimana yang ditampilkan pada Tabel 12 sebagai berikut:

Tabel 12. Hasil Klasifikasi Sistem

Kode	KD
SK1	3.2
SK2	3.5
SK3	3.2
SK4	3.1
SK5	3.9
SK6	3.1
SK7	3.3
SK8	3.2
SK9	3.2
SK10	3.2
SK11	3.2
SK12	3.3
SK13	3.7
SK14	3.7
SK15	3.5
SK16	3.7
SK17	3.3
SK18	3.3
SK19	3.3
SK20	3.2

Setelah dilakukan uji coba dengan menggunakan system, data tes soal diklasifikasi oleh tim ahli untuk dilakukan proses klasifikasi Kompetensi Dasar. Hasil klasifikasi menggunakan system dan hasil klasifikasi oleh tim ahli dibandingkan sehingga diperoleh akurasi dari proses klasifikasi yang dilakukan oleh system. Dengan mengacu pada Tabel Kompetensi Dasar Sistem Komputer pada Tabel 13 didapatkan hasil dari klasifikasi ahli adalah sebagai berikut:

Tabel 13. Hasil Klasifikasi Ahli

Kode	Acuan KD
SK1	3.2
SK2	3.5
SK3	3.2
SK4	3.1
SK5	3.9

Kode	Acuan KD
SK6	3.1
SK7	3.1
SK8	3.2
SK9	3.2
SK10	3.2
SK11	3.2
SK12	3.3
SK13	3.7
SK14	3.7
SK15	3.7
SK16	3.7
SK17	3.3
SK18	3.3
SK19	3.3
SK20	3.4

Tabel 14 Perbandingan Klasifikasi Ahli dengan Sistem Mapel Sistem Komputer

Kode	Acuan KD (Sistem)	Acuan KD (Ahli)
SK1	3.2	3.2
SK2	3.5	3.5
SK3	3.2	3.2
SK4	3.1	3.1
SK5	3.9	3.9
SK6	3.1	3.1
SK7	3.3	3.1
SK8	3.2	3.2
SK9	3.2	3.2
SK10	3.2	3.2
SK11	3.2	3.2
SK12	3.3	3.3
SK13	3.7	3.7
SK14	3.7	3.7
SK15	3.5	3.7
SK16	3.7	3.7
SK17	3.3	3.3
SK18	3.3	3.3
SK19	3.3	3.3
SK20	3.2	3.4

Dengan melihat hasil dari Tabel 14 terdapat kesesuaian hasil klasifikasi menggunakan system dengan klasifikasi ahli pada 17 soal yang telah diujikan. Sehingga akurasi dari hasil klasifikasi winnowing dengan nilai $n = 2$, $p = 2$, dan $w = 2$ pada mata pelajaran Sistem Komputer dengan data tes sebanyak 20 soal adalah $17/20 * 100\% = 85\%$.

Pada uji coba selanjutnya, dengan melakukan beberapa perubahan parameter, diperoleh hasil sebagaimana yang tampak pada Tabel 15 berikut:

Tabel 15. Perbandingan Klasifikasi Ahli dengan Sistem dengan $n = 3$, $w = 3$, $p = 2$

Kode	Acuan KD (Sistem)	Acuan KD (Ahli)
SK1	3.3	3.2
SK2	3.5	3.5
SK3	3.9	3.2
SK4	3.9	3.1
SK5	3.9	3.9
SK6	3.5	3.1
SK7	3.2	3.1
SK8	3.2	3.2

Kode	Acuan KD (Sistem)	Acuan KD (Ahli)
SK9	3.2	3.2
SK10	3.2	3.2
SK11	3.2	3.2
SK12	3.9	3.3
SK13	3.5	3.7
SK14	3.7	3.7
SK15	3.7	3.7
SK16	3.7	3.7
SK17	3.3	3.3
SK18	3.3	3.3
SK19	3.3	3.3
SK20	3.4	3.4

Dengan melihat hasil dari tabel 4.6 terdapat kesesuaian hasil klasifikasi menggunakan sistem dengan klasifikasi ahli pada 13 soal yang telah diujikan. Terdapat ketidak sesuaian hasil klasifikasi dari system dengan klasifikasi ahli pada soal dengan kode SK1, SK3, SK4, SK6. SK7, SK12 dan SK13 dengan total 7 soal yang tidak sesuai. Sehingga akurasi dari hasil klasifikasi *winnowing* dengan nilai $n = 3$, $p = 3$, dan $w = 2$ pada mata pelajaran Sistem Komputer dengan data tes sebanyak 20 soal adalah $13/20 * 100\% = 65\%$.

Pada percobaan ketiga dilakukan perubahan nilai $n = 4$, $w = 4$ dan $p = 2$ diperoleh hasil yang sama dengan percobaan sebelumnya, dapat dilihat pada Tabel 16 sebagai berikut:

Tabel 16 Perbandingan Klasifikasi Ahli dengan Sistem dengan $n = 4$, $w = 4$, $p = 2$

Kode	Acuan KD (Sistem)	Acuan KD (Ahli)
SK1	3.3	3.2
SK2	3.5	3.5
SK3	3.9	3.2
SK4	3.9	3.1
SK5	3.9	3.9
SK6	3.5	3.1
SK7	3.2	3.1
SK8	3.2	3.2
SK9	3.2	3.2
SK10	3.2	3.2
SK11	3.2	3.2
SK12	3.9	3.3
SK13	3.5	3.7
SK14	3.7	3.7
SK15	3.7	3.7
SK16	3.7	3.7
SK17	3.3	3.3
SK18	3.3	3.3
SK19	3.3	3.3
SK20	3.4	3.4

Dengan melihat hasil dari tabel 16 terdapat kesesuaian hasil klasifikasi menggunakan sistem dengan klasifikasi ahli pada 13 soal yang telah diujikan. Terdapat ketidaksesuaian hasil klasifikasi dari sistem dengan klasifikasi ahli pada soal dengan kode SK1, SK3, SK4, SK6. SK7, SK12 dan SK13 dengan total 7 soal yang tidak sesuai. Sehingga akurasi dari hasil klasifikasi *winnowing* dengan nilai $n = 4$, $p = 4$, dan $w = 2$ pada mata pelajaran Sistem Komputer dengan data tes sebanyak 20 soal adalah $13/20 * 100\% = 65\%$.

Dari tabel tersebut menunjukkan bahwa pada percobaan pertama dengan nilai parameter $n = 2$, $w = 2$ dan $p = 2$ diperoleh nilai akurasi klasifikasi soal oleh system sebesar 85%. Pada percobaan kedua, dengan data set yang sama dilakukan perubahan $n = 3$, $w = 3$ dan $p = 2$ dan diperoleh hasil akurasi

klasifikasi soal oleh sistem sebesar 65%. Pada percobaan ketiga, dengan melakukan perubahan $n = 4$, $w = 4$ dan $p = 2$ diperoleh hasil akurasi klasifikasi soal oleh sistem sebesar 65%.

V. KESIMPULAN DAN SARAN

Algoritma *Winnowing* merupakan algoritma yang digunakan dalam deteksi kesamaan menggunakan fungsi *hashing*. *Winnowing* adalah algoritma yang digunakan untuk melakukan proses *document fingerprinting*. *Fingerprint* yang diperoleh dijadikan dasar pembandingan kesamaan antara dua buah dokumen teks. Dari hasil penelitian ini, Algoritma *Winnowing* dapat digunakan untuk melakukan klasifikasi soal berdasarkan dengan Kompetensi Dasar yang relevan.

Dalam penelitian ini, dilakukan perubahan parameter n , w dan p dari Algoritma *Winnowing* untuk mengetahui kombinasi terbaik dari Algoritma *Winnowing*. Dari hasil penelitian diperoleh kombinasi terbaik dengan nilai $p=2$ jika nilai $n =2$ untuk proses N-gram dan $w = 2$ untuk proses *windowing* dengan nilai akurasi 85%.

Untuk penelitian selanjutnya diharapkan dilakukan penelitian lebih mendetail terhadap kombinasi parameter dari Algoritma *Winnowing* sehingga diperoleh akurasi yang lebih baik dalam melakukan klasifikasi antara soal dengan kompetensi dasar yang relevan.

DAFTAR PUSTAKA

- [1] Permendikbud, Nomor 24 Tahun 2016, “Kompetensi Inti dan Kompetensi Dasar Pelajaran Pada Kurikulum 2013 Pada Pendidikan Dasar dan Pendidikan Menengah”
- [2] K. Tung, N. Hung and L. Hauh, "A Comparison of Algorithms used to measure the Similarity between two documents", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 4, no. 4, pp. 1117-1121, 2015
- [3] Rifqi, Maharani., Shaufiah. 2011. Analisis dan Implementasi Klasifikasi Data Mining Menggunakan Jaringan Syaraf Tiruan dan Evolution Strategis. Institut Teknologi Telkom Bandung.
- [4] Suwanto Sanjaya, Ersad Alfarisy Absar, “Pengelompokan Dokumen menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour”, *Jurnal CoreIT*, Vol. 1, No.2, Desember, 2015
- [5] Keputusan Dirjen Dikdasmen, Nomor 330/D.D5/KEP/KR/2017, “Kompetensi Inti dan Kompetensi Dasar Mata Pelajaran Muatan Nasional (A), Muatan Kewilayahan (B), Dasar Bidang Keahlian (C1), Dasar Program Keahlian (C2), dan Kompetensi Keahlian (C3)”.
- [6] Agung Toto Wibowo, Kadek W. Sudarmadi, Ari M. Barmawi, “*Comparision Between Fingerprint and Winnowing Algorithm to Detect Plagiarism Fraud on Bahasa Indonesia Documents*”, *International Conference of Information and Communication Technology (ICoICT)*, 2013
- [7] Schleimer Saul, Wilkerson Daniel S., Aiken Alex, "Winnowing: Local Algorithms for Document Fingerprinting," in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, NY, USA, 2003, pp. 76-85
- [8] A. A. P. Ratna, dkk, “ Latent Semantic Analysis and Winnowing Algorithm Based Automatic Japanese Short Essay Answer Grading System Comparative Performance”, *IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, 2019
- [9] Jarwati et al, “Penerapan Algoritma *Winnowing* Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)”, *Berkala Saintek* 2017, V (1) : 11-20
- [10] Dedi Leman, dkk, “Rabin Karp And Winnowing Algorithm For Statistic of Text Document Plagiarism Detection”, in *The 7th International Conference on Cyber and IT Service Management (CITSM 2019)*
- [11] Eric Ganiwijaya Hasan, dkk, “The Implementation of Winnowing Algorithm for Plagiarism Detection in Moodle-based E-Learning”, *ICIS*, June 6-8, 2018, Singapore
- [12] A. A. P. Ratna, D. L. Luhurkinanti, I. Ibrahim, D. Husna, and P. D. Purnamasari, “*Automatic Essay Grading System for Japanese Language Examination Using Winnowing Algorithm*,” in *2018 International Seminar on Application for Technology of Information and Communication*, 2018, pp. 565–569