

## Penggunaan N-Gram dan Jaro Winkler Distance pada Aplikasi Kelas Daring untuk Deteksi Plagiat

Lukman Hakim<sup>1\*</sup>

<sup>1</sup>Teknik Informatika, Universitas Bunda Mulia, Jakarta Utara, Jl. Lodan raya No 2, 14430

\*Corresponding Author : lhakim2710@gmail.com

### Abstrak

Perkembangan teknologi *e-learning* yang diterapkan oleh berbagai perguruan tinggi memberikan kemudahan tanpa perlu bertatap muka secara langsung, hal ini tentu juga banyak dampak negatif dan positif salah satu dampak negatif apakah mahasiswa dapat terkontrol dengan baik seperti tugas yang mahasiswa kerjakan tidak plagiat terhadap tugas mahasiswa lainnya, hal ini perlu dibuat aplikasi yang berfungsi untuk memberikan kemudahan dosen untuk melacak tingkat kesamaan tugas antara mahasiswa satu dengan mahasiswa lainnya, dengan perancangan aplikasi deteksi plagiat menggunakan algoritma N-gram dan *Jaro Winkler Distance* dapat melakukan pengecekan kesamaan kata pada tugas forum mahasiswa pada kelas daring, dari pengujian 18 tugas forum pada algoritma N-Gram dan jaro winkler terdapat kesamaan kata rata-rata sebesar 75,56%. Pengujian secara acak pada materi acuan secara otomatis belum dilakukan.

**Kata kunci:** *e-learning*, N-gram, jaro winkler distance, plagiat

### Abstract

*The development of e-learning technology that is applied by various tertiary institutions provides convenience without the need for face-to-face meeting, this is of course also many negative and positive impacts, one of the negative impacts is whether students can be controlled properly such as assignments that students do not plagiarist to other student assignments , this needs to be made an application that serves to make it easier for lecturers to track the level of similarity of tasks between one student and another student, by designing a plagiarism detection application using the N-gram algorithm and Jaro Winkler Distance can check the similarity of words in the student forum assignments in online classes , from testing 18 forum assignments on the N-Gram algorithm and Jaro Winkler there are similarities in words by an average of 75.56%. Random testing of reference material automatically has not been carried out.*

**Keywords :** *e-learning*, N-gram, jaro winkler distance, plagiarism.

### PENDAHULUAN

Pemanfaatan teknologi informasi dalam hal ini *e-learning* diperlukan tidak hanya pendidik yang terampil memanfaatkan teknologi serta teknologi untuk pembuatan bahan ajar, akan tetapi diperlukan suatu rancangan agar dapat melaksanakan pembelajaran dengan efektif (Hanum, 2013). *E-learning* sebagai pendidikan terencana yang menggunakan media elektronik, yang mencakup pembelajaran jarak jauh melalui Internet. *e-learning* identik dengan pembelajaran daring. Lingkungan belajar hibrida hanyalah *e-learning* yang menggunakan pendidikan tradisional tatap muka (FtF)

bersama dengan media elektronik (Glancy, 2013).

*E-Learning* merupakan cara pembelajaran yang bertujuan untuk memberikan kemudahan dan menggunakan teknologi informasi dalam proses pengajaran (Fuad, 2013). Komponen tugas forum merupakan bagian dari penilaian keaktifan dan kehadiran mahasiswa pada matakuliah *e-learning*, untuk memenuhi syarat tersebut mahasiswa melakukan tindakan salin jawaban mahasiswa atau melakukan plagiat.

Berdasarkan kamus besar bahasa Indonesia plagiat adalah pengambilan karangan (pendapat

dan sebagainya) orang lain dan menjadikannya seolah-olah karangan (pendapat dan sebagainya) sendiri, misalnya menerbitkan karya tulis orang lain atas nama dirinya sendiri (KBBI, 2019). Kecurangan ini harus dihilangkan demi meningkatkan kemampuan mahasiswa dalam belajar. Tujuan penelitian deteksi plagiat tugas forum mahasiswa kelas daring mengetahui sejauh mana plagiat mahasiswa mengerjakan tugas forum pada kelas daring. **Plagiarisme** yang terkadang disebut plagiat menurut Cosma dan Joy (2008) sering dinyatakan sebagai menyalin pekerjaan orang lain (sebagai contoh dari siswa lain atau dari sumber-sumber seperti buku teks), dan lalai untuk memberikan pengakuan yang tepat dari sumber (pencetus bahan yang ditiru). (Lisangan, 2013).

Berdasarkan hal tersebut ditemukan banyak sekali mahasiswa menyalin jawaban mahasiswa lainnya, perlu dibuat mekanisme cara mendeteksi tingkat plagiarisme mahasiswa dalam mengerjakan forum *e-learning* dan meminimalisir kecurangan, dengan metode N-Gram dapat membandingkan setiap teks dokumen dengan dokumen lainnya, dalam indikasi kesamaan, N-Gram membaca dokumen yang menjadi masukan kemudian menghitung jumlah kemunculan setiap N-Gram pada dokumen tersebut (Badrus Zaman, 2015). Algoritma *Jaro Winkler* adalah algoritma untuk mengukur kesamaan antara dua string (Panji Novantara, 2018). Ada beberapa tipe plagiat yaitu (Panji Novantara, 2018) :

1. *Word for word* adalah menyalin kata perkata tanpa merubah sedikit kalimat.
2. *Plagiat of authorship* adalah mengakui karya orang lain sebagai karyanya sendiri, serta tidak mencantumkan penulis aslinya.
3. Plagiat of ide adalah mengakui hasil pemikiran orang lain.

Kelebihan algoritma *jaro winkler distance* memiliki 3 bagian yaitu (Winkler., 2006) :

1. Menghitung panjang *string*
2. Menemukan jumlah karakter yang sama didalam dua string, dan
3. Menemukan jumlah transposisi

Penelitian ini diharapkan dapat mengukur tingkat kemandirian mahasiswa pada kelas *e-learning* Universitas XYZ sebagai media pembelajaran yang efektif berdasarkan *framework* yaitu dimensi pedagogi.

## METODE

1. Pengumpulan data berupa dokumen tugas forum mahasiswa pada matakuliah *e-learning* dengan mengambil beberapa contoh dokumen jawaban tugas forum serta melakukan pengamatan dengan cara memfilter dokumen yang memiliki kesamaan dan mengumpulkan beberapa literatur yang berhubungan dengan penelitian
2. Melakukan pengujian metode N-Gram dan *Jaro Winkler* terhadap data berupa tugas mahasiswa kelas *e-learning* atau matakuliah *e-learning*.
3. Melakukan pengecekan *similarity* pada kesamaan data

Teknik Analisis data

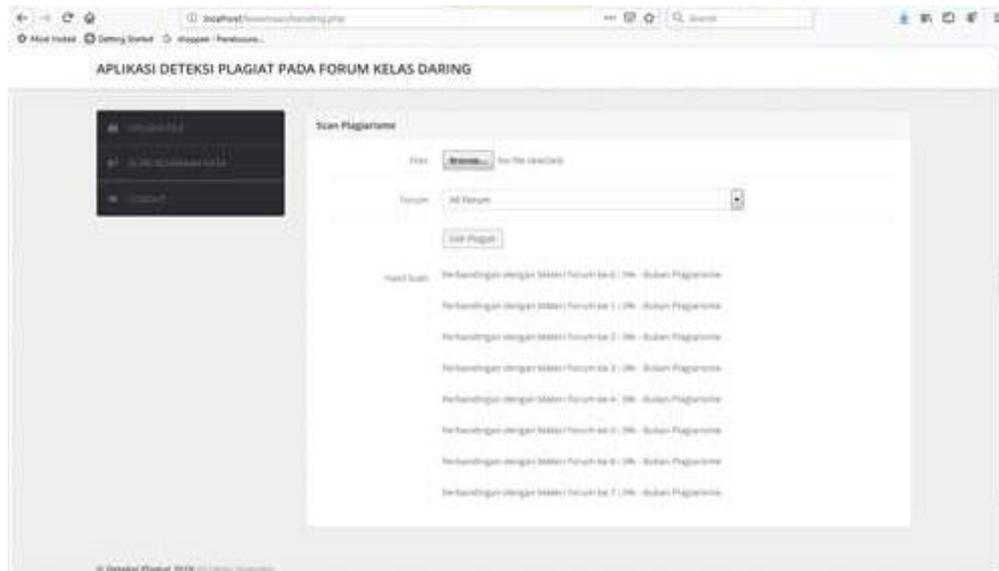
Analisis data pada penelitian ini adalah :

- a. Mendokumentasikan semua data yang berhubungan dengan tugas forum mahasiswa pada kelas *e-learning*.
- b. Melakukan perencanaan dengan menentukan lingkup dan penggunaan bahasa pemrograman yang ideal serta algoritma yang sesuai.
- c. Pelaksanaan dengan membuat program prototype untuk menguji sejauhmana algoritma efektif.
- d. Evaluasi, melakukan pengujian 18 tugas forum mahasiswa dengan algoritma Ngram dan jaro winkler.

## HASIL DAN PEMBAHASAN

### Aplikasi Deteksi Plagiat tugas Kelas daring

Penggunaan aplikasi ini untuk mengukur sejauhmana kemandirian mahasiswa mengerjakan tugas forum kelas daring tanpa melakukan tindakan plagiat, dapat dilihat pada gambar 1. Tampilan aplikasi deteksi plagiat.

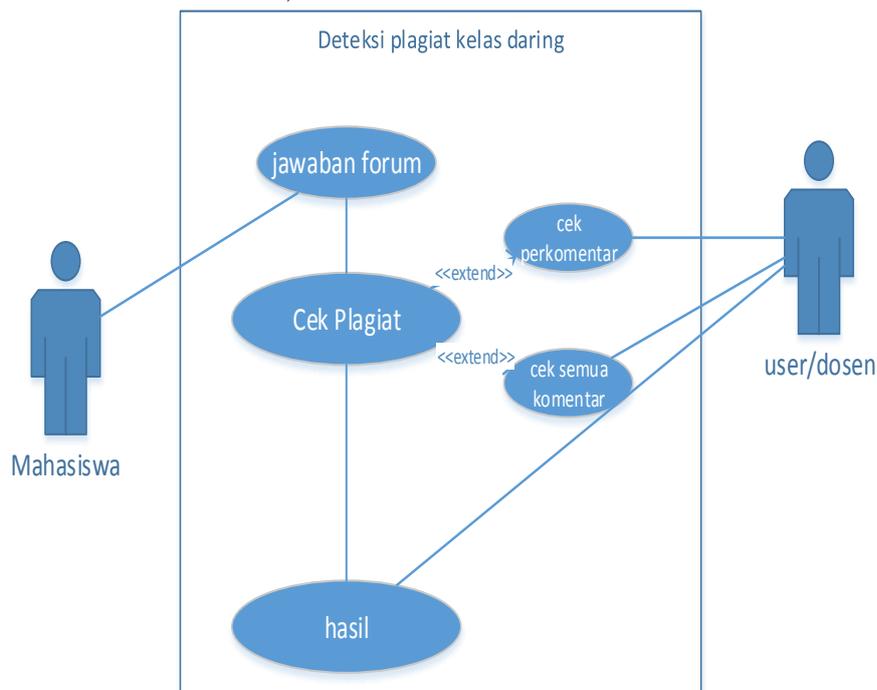


Gambar 1. Aplikasi Deteksi plagiat tugas forum kelas daring

**Usecase Diagram Deteksi plagiat kelas daring**

Pada gambar 2. Perancangan usecase diagram mahasiswa mengirimkan jawaban forum berdasarkan soal yang sudah ditentukan, lalu sistem melakukan pengecekan plagiat pada jawaban forum mengacu pada jawaban pertama yang dikirimkan oleh mahasiswa, untuk

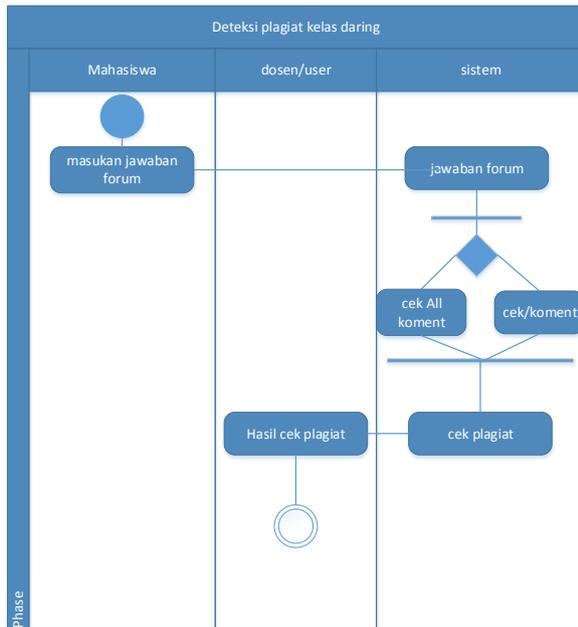
pengecekan bisa dilakukan perjawaban forum atau secara keseluruhan, hasil dapat diketahui dosen sejauhmana tingkat kesamaan, ambang batas 0-20% masih dianggap tidak melakukan plagiat.



Gambar 2. Usecase diagram deteksi plagiat kelas daring

**Aktivity diagram Deteksi plagiat kelas daring**

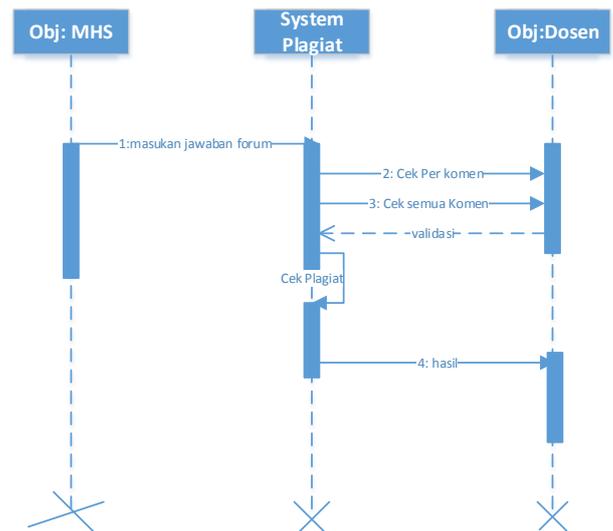
Pada gambar 3. Diagram activity menggambarkan urutan aktivitas dari pengguna, dengan adanya aplikasi deteksi plagiat memberikan penilaian dan kemandirian mahasiswa untuk berpikir kreatif tanpa perlu menyalin jawaban teman mereka, apabila ada indikasi terjadinya plagiat dapat mempengaruhi penilaian.



Gambar 3. Activity diagram deteksi plagiat

**Diagram Sequence cek Plagiat**

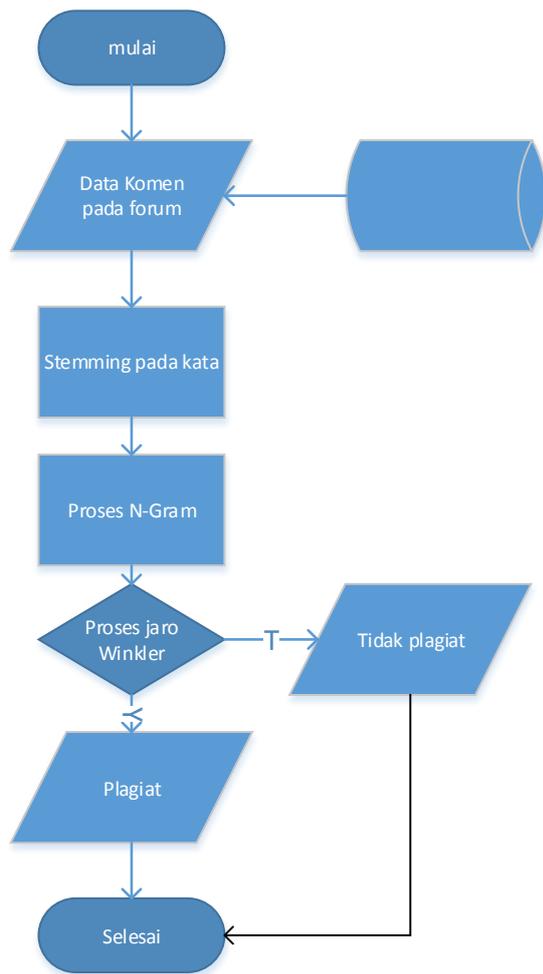
Diagram sequence menggambarkan interaksi antar objek, pada konsep objek oriented setiap objek yang memiliki keterkaitan, objek sistem yang berfungsi melakukan deteksi pada setiap koment yang masuk dan hasil langsung dapat dilihat pada gambar 4. Sequence diagram deteksi plagiat kelas daring sebagi berikut.



Gambar 4. Sequence Diagram deteksi plagiat kelas daring

**Proses penerapan algoritma N-Gram dan Jaro Winkler**

Penggunaan algoritma N-Gram dan *Jaro Winkler Distance* untuk mengetahui sejauh mana algoritma ini dapat diterapkan dalam deteksi kesamaan kata. Data yang masuk pada kelas daring dilakukan *stemming* perkata dengan dibanding dengan koment acuan, setelah itu algoritma *jaro winkler* melakukan pengecekan dan menentukan kriteria berdasarkan nilai plagiat yang sudah ditentukan, apabila total kesamaan lebih dari 20% maka nilai dari forum melakukan plagiat, untuk gambar 5. Dapat dilihat *flowchar* proses N-gram dan *jarok winkler* sebagai berikut.



Gambar 5. Flowchar proses algoritma

**Pengujian aplikasi**

Untuk mengetahui sejauhmana algoritma N-gram dan Jaro Winkler ini dapat bekerja dilakukan pengujian dengan 18 tugas forum mahasiswa secara acak dapat dilihat pada tabel 1. Pengujian aplikasi terbagi 2 pengujian setiap pengujian terbagi 8 komentar yang di uji , dimana untuk setiap deteksi kesamaan komentar pertama kali masuk forum dijadikan acuan untuk komentar lainnya, sebagai berikut.

Tabel 1. Pengujian Aplikasi

No	Mahasiswa	Jawaban Forum	Presentase kemiripan
1	Andreas	Dampak negatif hiburan seperti games pada anak, yaitu : 1. Merasa gelisah dan mudah marah apabila tidak diijinkan bermain.	100%

		<p>2. Yang ada di pikirannya hanyalah permainan online, sehingga ketika ibu mengajaknya ngobrol, topik yang ia bicarakan juga tentang permainan tersebut.</p> <p>3. Selalu ingin main game terus menerus dan susah bila disuruh berhenti.</p> <p>4. Tidak peduli dengan orang-orang di sekitar, bahkan enggan bersosialisasi dengan teman-teman sebayanya.</p> <p>5. Mengalami migrain dan mata lelah karena terlalu lama memandang layar komputer atau gadget.</p>	
2	Berin	<p>Dampak negatif games untuk anak - anak adalah lupa waktu dimana ketika sang anak sudah masuk ke dalam dunia bermain anak - anak sering kali melupakan waktu untuk melakukan kegiatan lain seperti makan dan pergi ke toilet dan atau belajar.</p> <p>#janganganguwaktumainde nganbelaja</p>	78,85%
3	Jonathan	<p>Kesehatan mata terganggu Karena anak terlalu memfokuskan mata nya ke layar komputer atau gadget membuat kesehatan mata berkurang, seperti mata lelah, bahkan bisa mengalami kerusakan saraf mata.</p> <p>- Gangguan motorik Pada saat bermain game, anak cenderung hanya duduk saja dan tidak menggerakkan badannya pada saat bermain. Akibatnya kemampuan motorik akan menurun, sehingga pertumbuhan anak tidak maksimal dan berisiko mengalami obesitas karena lemak nya menumpuk.</p> <p>- Nyeri sendi Di beberapa kondisi, terkadang anak lebih memilih untuk bermain game dengan posisi membungkuk atau tiduran, Dan posisi tersebut bukanlah posisi duduk yang sehat. Jika tetap di teruskan, maka dapat membuat otot nya terasa kaku dan dapat mengalami nyeri sendi.</p> <p>- Menurunkan tingkat konsentrasi anak</p>	66,41%

		Ketika anak senang bermain game, akan terjadi perubahan pada struktur dendrit sel-sel di dalam otaknya. Hal ini mengakibatkan konsentrasi anak menurun, sehingga ia mudah lupa dan gagal fokus. Paparan radiasi dari perangkat elektronik juga bisa melemahkan konsentrasi anak	
4	Jessica	Dampak negatif dari games dan bentuk hiburan lain adalah A. Menurunkan niat anak dalam belajar. (Anak lebih cenderung mencari games dibanding belajar karena itu menyenangkan) B. Mengurangi kemampuan mata (warna yang cerah dari aplikasi tersebut akan mengurangi ketajaman penglihatan mata jika melihat secara terus menerus.) C. Adiksi Games. (Seringkali anak menjadi keranjingan main games sampai lupa waktu. Hal ini bahkan memunculkan kasus dimana anak melupakan hal yang harus dikerjakan bahkan sampai sakit ketika dikarang main games.)	86,66%
5	Chandra	Dampak negatif untuk anak-anak adalah membuat anak tersebut menjadi tidak disiplin karena anak-anak cenderung melupakan waktu saat bermain, sedangkan disiplin waktu adalah salah satu hal yang harus diajarkan sejak kecil, kalau tidak kebiasaan tersebut akan terbawa hingga dewasa seperti sering tidak tepat waktu dalam melakukan berbagai hal.	81,5%
6	Ryan	Dampak negatif hiburan seperti games pada anak-anak, yaitu dapat memberikan dampak kecanduan games, jika anak-anak kecanduan bermain games maka akan membuat mereka malas belajar dan tidak ingin mengerjakan Pekerjaan Rumah mereka, anak-anak tersebut akan lebih memilih bermain games dibanding belajar. Selain itu dapat mengganggu kesehatan mereka karena	85,27%
		terlalu lama bermain games misalnya dapat menyebabkan sakit mata atau kram tangan.	
7	Mikael	Dampak negatif hiburan seperti games pada anak-anak adalah a) Membuat kecanduan yang berlebihan. Bermain game tentunya dapat membuat kecanduan karena game merupakan sebuah hiburan. Tapi bermain game secara terus-menerus dapat membuat anak-anak lupa waktu. Hal ini membuat mereka lupa dengan pekerjaan, makan, dan lain-lain. b) Membuat mata rusak. Menatap layar monitor berjam-jam untuk bermain game tentunya akan membuat mata anak-anak berair atau perih. Bahkan jika terlalu sering, mata anak-anak bisa minus dan harus menggunakan kacamata. c) Membuat malas. Bermain game akan membuat anak-anak lupa segalanya. Mereka hanya fokus terhadap game yang dimainkan. Anak-anak akan menjadi malas bekerja atau mengerjakan tugas. d) Membuat tidak konsentrasi. Saat anak-anak belajar atau bekerja, mereka tidak konsentrasi terhadap apa yang sedang dikerjakan. Pikirannya hanya tertuju pada game-game yang telah mainkan.	80,99%
8	Putri	Menurut saya, dampak negatif dari hiburan seperti game terhadap anak-anak terbagi menjadi beberapa aspek seperti aspek sosial, aspek mental/kepribadian, aspek pendidikan. Aspek sosial: anak-anak yang kecanduan bermain game cenderung sulit berkomunikasi dengan orang lain, hal ini dikarenakan dengan bermain game anak-anak akan kurang melakukan komunikasi dengan orang lain di sekitarnya Aspek mental/kepribadian: anak-anak cenderung	65,79%





		Menatap layar monitor berjam-jam untuk bermain game tentunya akan membuat mata anak-anak berair atau perih. Bahkan jika terlalu sering, mata anak-anak bisa minus dan harus menggunakan kacamata. c) Membuat malas. Bermain game akan membuat anak-anak lupa segalanya. Mereka hanya fokus terhadap game yang dimainkan. Anak-anak akan menjadi malas bekerja atau mengerjakan tugas. d) Membuat tidak konsentrasi. Saat anak-anak belajar atau bekerja, mereka tidak konsentrasi terhadap apa yang sedang dikerjakan. Pikirannya hanya tertuju pada game-game yang telah mainkan.	
17	David	Menurut saya, dampak negatif dari hiburan seperti game terhadap anak-anak terbagi menjadi beberapa aspek seperti aspek sosial, aspek mental/kepribadian, aspek pendidikan. Aspek sosial: anak-anak yang kecanduan bermain game cenderung sulit berkomunikasi dengan orang lain, hal ini dikarenakan dengan bermain game anak-anak akan kurang melakukan komunikasi dengan orang lain di sekitarnya. Aspek mental/kepribadian: anak-anak cenderung mengikuti apa yang ada di dalam game, seperti kekerasan, kata kasar, sampai karakter atau sifat dari karakter dalam game. Aspek pendidikan: anak-anak yang kecanduan dalam game kebanyakan menghabiskan waktu untuk bermain game dan tidak belajar sehingga mengakibatkan tingkat kecerdasan yang rendah, hal ini mengakibatkan timbulnya beberapa permasalahan di dalam game itu sendiri dimana player dengan tingkat kecerdasan yang rendah menjadi racun sendiri dalam	80,99%

		game tersebut, sebagai contoh salah satu game online yang saya mainkan, terdapat kelompok-kelompok pemain dengan kecerdasan rendah yang disebut "DOMBA"	
18	Lina	menurut saya dampak negatif hiburan seperti games pada anak-anak adalah <ul style="list-style-type: none"> <li>• kecanduan</li> </ul> seringnya anak-anak dalam bermain game akan menimbulkan kecanduan. maksud dari kecanduan disini ada ketergantungan, anak-anak akan semakin sering bermain game dan tidak peduli pada lingkungan sekitarnya. <ul style="list-style-type: none"> <li>• malas</li> </ul> selain kecanduan, anak-anak juga akan menjadi malas. tidak mau belajar karena terlalu seru bermain game. <ul style="list-style-type: none"> <li>• tidak disiplin</li> </ul> disini adalah maksudnya itu anak-anak jadi susah diatur oleh orang tuanya	65,79%

Berdasarkan pengujian dari 18 tugas dengan mengacu pada tugas 1 dan 10 dapat dirata-ratakan keberhasilan penggunaan algoritma N-gram 79,56%.

#### **Pengujian Kesamaan kalimat berdasarkan manual**

Berdasarkan pengujian tabel 2. Kesamaan kata pada algoritma Jaro Winkler, setiap penerapan algoritma *Jaro Winkler Distance* dan Ngram belum sepenuhnya bisa membuktikan identifikasi plagiat. Karena pengujian kedua algoritma ini sebatas pengujian kata bukan kesamaan makna.

Tabel 2. Pengujian Kesamaan kata

No	Jumlah Kata Acuan	Jumlah Kata	Persentase Kesamaan
1	79	150	66.41%
2	79	45	78.85%
3	79	76	86.66%
4	79	47	81.5%
5	79	59	85.27%
6	79	125	80.99%
7	79	68	65.79%
8	79	145	80.99%
9	79	150	78.85%
10	79	71	66.41%
11	79	198	66.42%
12	79	47	81.5%
13	79	76	86.66%
14	79	150	66.41%
15	79	59	81.5%
16	79	68	65.79%
18	79	150	78.85%
Rata2			76,32%

### SIMPULAN DAN SARAN

Hasil pengujian pada algoritma N-gram dan *Jaro Winkler Distance* pada aplikasi deteksi plagiat tugas forum kelas daring dari 18 tugas forum dengan rata-rata 76,32%. Untuk aplikasi ini belum bisa melakukan pengujian secara berkala pada setiap tugas forum mahasiswa pada kesamaan tugas. Penggunaan N-gram dan *Jaro Winkler* belum bisa mendeteksi kesamaan makna pada setiap kalimat hal tersebut berdasarkan pengujian manual. Dengan adanya penambahan fitur deteksi kesamaan kata pada kelas daring memberikan kemudahan dosen untuk menilai mahasiswa dalam kemandirian mengerjakan tugas secara baik.

### SARAN

Untuk penelitian berikutnya sebaiknya menggunakan algoritma yang dapat mengenal makna berdasarkan kalimat, sehingga deteksi kesamaan kalimat atau plagiat lebih baik.

### DAFTAR PUSTAKA

- Badrus Zaman, E. H. 2015. Sistem Deteksi Bahasa pada Dokumen. *Jurnal Multinetics Vol. 1 No. 2 November 2015*, 21-26.
- Fuad, H. 2013. Rancang Bangun Sistem Informasi E-Learning berbasis web di SMK 1 Tangerang. *JURNAL SISFOTEK GLOBAL, Vol 3*, 4-8.
- Glancy, F. H. 2013. A Conceptual Learner-Centered e-Learning Framework. *Journal of Higher Education Theory and Practice*, 13.
- Hanum, N. S. 2013. The Effectiveness Of E-Learning As Instructional Media (Evaluation Study Of E-Learning Instructional Model Insmk Telkom Sandhy Putra Purwokerto). *Jurnal Pendidikan Vokasi, Vol 3, nomor 1, februari 2013*, 90-102.
- Lisangan, E. A. 2013. Implementasi N-Gram Technique Dalam Deteksi Plagiarisme pada tugas mahasiswa. *Jurnal Tematika Vol 1 No.2*, 24-29.
- Panji Novantara, O. P. 2018. Implementasi Algoritma Jaro-Winkler Distance Untuk Sistem Pendeteksi Plagiarisme Pada Dokumen Skripsi. *Jurnal Buffer Informatika Vol 3 No.2*
- Retrieved 5 15, 2019, from KBBI: <https://kbbi.kemdikbud.go.id/>
- Winkler., W. E. 2006. *Overview of record linkage and current research directions*. RRS: Research Report Series.